THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Molecular mechanisms regulating *N*-glycosylation site selection by yeast oligosaccharyltransferase: Role of Ost3p and Ost6p

Muhammad Fairuz Bin Jamaluddin

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2014

School of Chemistry and Molecular Biosciences

Abstract

Asparagine (*N*)-linked glycosylation is a common post-translational modification that regulates the structure and function of secretory and membrane proteins in eukaryotes. This process is catalysed by the multiprotein complex oligosaccharyltransferase (OTase), which glycosylates selected asparagine residues in nascent polypeptides as they enter the endoplasmic reticulum (ER) through the translocon. *N*-glycosylation is crucial in protein folding, and the precise glycans present on mature glycoproteins are often vital for their biological functions. Two isoforms of OTase exist in Bakers' yeast *Saccharomyces cerevisiae*, determined by the presence of either of the homologous proteins Ost3p or Ost6p. Previous studies have identified that efficient glycosylation of some specific sites requires Ost3p, while others require Ost6p. Ost3p and Ost6p have thioredoxin-like ER soluble domains with grooves adjacent to their redox active sites, and our recent *in vitro* assays have shown that Ost3p and Ost6p can bind stretches of hydrophobic polypeptide with complementary characteristics to these peptide-binding grooves. A model of Ost3/6p function has been proposed based on this *in vivo* and structural data in which we hypothesize that Ost3p and Ost6p transiently bind nascent polypeptide to increase the glycosylation efficiency of selected asparagines.

In the first part of the thesis, we tested a key prediction of this model using yeast genetics and glycoproteomics and established that this peptide-binding activity of Ost3/6p is physiologically relevant in *N*-glycosylation *in vivo*, suggesting that nascent polypeptide binds directly to the grooves in Ost3/6p. Further, we identified the precise sites of interaction between the Gas1p substrate nascent polypeptide and the Ost3/6p peptide binding groove *in vivo*, using a combination of site-directed mutagenesis and mass spectrometry glycoproteomics. Finally, using our *in vitro* peptide binding assay, we demonstrated and validated that a stretch of Gas1p we identify as interacting with Ost3/6p *in vivo* also

interacted with Ost6p *in vitro*. Together, this study advances our understanding of Ost3/6p function in which they transiently bind stretches of nascent polypeptide substrate proximal and *N*-terminal to specific glycosylation sites, which inhibits local protein folding and thereby increases glycosylation efficiency.

In the second part of the thesis, we investigated the *in vitro* interactions between substrate polypeptide and Ost3/6p by peptide affinity chromatography and Selected Reaction Monitoring (SRM) Mass Spectrometry. We developed SRM-MS for improved relative quantification and sensitivity for detection in conjunction with peptide affinity chromatography. We show that Ost3p and Ost6p bind distinct subsets of peptides from a yeast cell wall glycoprotein substrate, with Ost3p binding peptides rich in aromatic acids and Ost6p binding peptides rich in aliphatic amino acids. Our data supports a model of Ost3/6p function in which they transiently bind complementary hydrophobic stretches of nascent polypeptide substrate to optimally inhibit protein folding, thereby increasing glycosylation efficiency at nearby asparagine residues.

In the third part of the thesis, we developed another improved analytical method for site-specific glycosylation analysis using PNGase F to label glycosylation sites with an asparagines-aspartate conversion that creates a new endoproteinase AspN cleavage site, followed by proteolytic digestion, and detection of peptides and glycopeptides by liquid chromatography electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS). This approach will be useful in site-specific glycosylation analysis in many model systems and clinical application. Overall, these two novel analytical methods introduced in the second and the third part of this thesis will provide advanced knowledge for improved detection and quantification of glycosylation sites *in vivo* and peptide interactions *in vitro*.

In the fourth part of the thesis, we used a novel strategy for engineering glycoprotein stability to determine the detailed function of Ost3p and Ost6p *in vitro*. The ER lumenal domain of Ost3p has a single glycosylation sequon, while Ost6p has no glycosylation sequons and contains a cluster of acidic amino acids at the position corresponding to the Ost3p sequon. We showed that incorporation of targeted like-charged amino acids at *N*-glycosylation sequons in Ost3/6p increases *in vitro* stability and activity. Optimal stabilization through minimal number of local point mutations in Ost3p glycosylation sequon showed increased *in vitro* binding of peptides and these peptides are complementary to the characteristics of the peptide binding groove of Ost3, a key determinant of transient nascent polypeptide tethering by Ost3p, consistent with the function of Ost3p and Ost6p in modulating *N*-glycosylation substrate selection by OTase activity *in vivo*.

In summary, this thesis has made a significant contribution towards our understanding the molecular mechanisms of glycosylation site selection by OTase activity, the connectivity between glycoprotein folding and *N*-glycosylation in the endoplasmic reticulum, and opens up potential applications in synthetic glycobiology such as engineering of glycosylation sites and OTases. Our data provides advance information towards solving the model of the mechanisms by which Ost3/6p increase the efficiency of site-specific *N*-glycosylation. Here we detailed the mechanism of this regulatory system through mapping the precise sites of interactions between Ost3/6p and Gas1p nascent polypeptide which was not previously identified. This is a novel insight into the mechanisms of co-translocational polypeptide modification in the endoplasmic reticulum.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the General Award Rules of The University of Queensland, immediately made available for research and study in accordance with the *Copyright Act 1968*.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

1. **Peer-reviewed papers**

Bailey UM, **Jamaluddin MF**, Schulz BL. (2012). Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J Proteome Res*. 11(11):5376-5383. (Incorporated as Chapter 4)

Mohd Yusuf SN, Bailey UM, Tan NY, **Jamaluddin MF**, Schulz BL. (2013). Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem Biophys Res Commun*. 432(3):438-443.

Tan NY, Bailey UM, **Jamaluddin MF**, Mahmud SH, Raman SC, Schulz BL. (2014). Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun*. 5: 3099. (Incorporated as Chapter 5)

Jamaluddin MF, Bailey UM, Schulz BL (2014). Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation. *Mol Cell Proteomics*. 13(12):3286-3293. (Incorporated as Chapter 2)

Jamaluddin MF, Bailey UM, Schulz BL (2014). Selected reaction monitoring Mass Spectrometry for targeted peptide mult-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p. In Progress. (Incorporated as Chapter 3)

2. Conference abstracts

Jamaluddin MF, Bailey UM, Tan NY, Schulz BL. Identification of precise sites of interaction between the Gas1p substrate to OTase subunits Ost3/6p peptide binding groove in *N*-glycosylation in vivo. *8th Annual School of Chemistry and Molecular Biosciences Research Students Symposium,* November 2012, University of Queensland, Brisbane, Australia – *Poster*.

Jamaluddin MF, Bailey UM, Schulz BL. Ost3p and Ost6p sequester nascent substrate polypeptide to locally enhance *N*-glycosylation by oligosaccharyltransferase. *International Postgraduate Symposium for the School of Biomedical Sciences*, 2013, University of Queensland, Brisbane, Australia – *Poster*.

Jamaluddin MF, Bailey UM, Schulz BL. Ost3p and Ost6p sequester nascent substrate polypeptide to locally enhance *N*-glycosylation by oligosaccharyltransferase. *First Biennial School of Chemistry and Molecular Biosciences Research Symposium.* 2013, Translational Research Institute, Brisbane, Australia – *Poster*.

Jamaluddin MF, Bailey UM, Schulz BL. Ost3p and Ost6p sequester nascent substrate polypeptide to locally enhance *N*-glycosylation by oligosaccharyltransferase. *9th Annual School of Chemistry and Molecular Biosciences Research Students Symposium,* Novmeber 2013, University of Queensland, Brisbane, Australia – *Oral Presentation*

Publications included in this thesis

Jamaluddin MF, Bailey UM, Schulz BL (2014). Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation. *Mol Cell Proteomics*. 13(12):3286-3293 (Incorporated as Chapter 2)

Contributor	**Statement of contribution**
Jamaluddin, MF (Candidate)	Designed experiments (80%) Wrote the initial draft (100%) Performed *in vivo* yeast experiments (85%) Performed *in vitro* peptide binding assays (100%) Performed mass spectrometry* Statistical analysis of data (50%) Made figures and tables (70%) Wrote and edited paper *
Bailey, UM	Designed experiments* Performed *in vivo* yeast experiments (15%) Performed mass spectrometry* Wrote and edited paper*
Schulz, BL	Designed experiments* Performed mass spectrometry* Mass spectrometry analysis* Made figures and tables (30%) Statistical analysis of data (50%) Wrote and edited paper*

*All authors contributed equally

Jamaluddin MF, Bailey UM, Schulz BL (2014). Selected reaction monitoring Mass Spectrometry for targeted peptide mult-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p. In Progress. (Incorporated as Chapter 3)

Contributor	**Statement of contribution**
Jamaluddin, MF (Candidate)	Designed experiments* Wrote the initial draft (100%) Performed *in vitro* peptide binding assays (100%) Performed mass spectrometry* Statistical analysis of data (60%) Made figures and tables (50%) Wrote and edited paper *
Bailey, UM	Designed experiments* Performed mass spectrometry* Wrote and edited paper*
Schulz, BL	Designed experiments* Performed mass spectrometry* Mass spectrometry analysis* Made figures and tables (50%) Statistical analysis of data (40%) Wrote and edited paper*

*All authors contributed equally

Bailey UM, **Jamaluddin MF**, Schulz BL. (2012). Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J Proteome Res*. 11(11):5376-5383. (Incorporated as Chapter 4)

Contributor	**Statement of contribution**
Bailey, UM	Designed experiments* Wrote the initial draft (100%) Performed *in vivo* yeast experiments (80%)

	Performed mass spectrometry* Statistical analysis of data (60%) Made figures and tables (100%) Wrote and edited paper *
Jamaluddin, MF (Candidate)	Designed experiments* Performed *in vivo* yeast experiments (20%) Wrote and edited paper*
Schulz, BL	Designed experiments* Performed mass spectrometry* Mass spectrometry analysis (50%) Statistical analysis of data (40%) Wrote and edited paper*

*All authors contributed equally

Tan NY, Bailey UM, **Jamaluddin MF**, Mahmud SH, Raman SC, Schulz BL. (2014). Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun*. 5:3099. (Incorporated as Chapter 5)

Contributor	**Statement of contribution**
Tan, NY	Designed experiments* Wrote the initial draft (80%) Performed sequence analysis (40%) Characterised the proteins (40%) Statistical analysis of data (50%) Made figures and tables (30%) Wrote and edited paper *
Bailey, UM	Designed experiments* Performed the growth inhibition assay Statistical analysis of data (25%) Made figures and tables (30%) Wrote and edited paper *
Jamaluddin, MF (Candidate)	Designed experiments* Performed *in vitro* peptide binding assays

	(100%) Performed mass spectrometry (60%) Wrote and edited paper *
Mahmud, SH	Designed experiments* Characterised the proteins (30%) Performed *in vivo* growth yeast assays Wrote and edited paper*
Raman, SC	Designed experiments* Performed sequence analysis (30%) Wrote and edited paper*
Schulz, BL	Designed experiments* Performed sequence analysis (30%) Characterised the proteins (30%) Statistical analysis of data (25%) Performed mass spectrometry (40%) Mass spectrometry analysis Made figures and tables (40%) Wrote and edited paper*

*All authors contributed equally

Contributions by others to the thesis

Contributions by others to the publications incorporated into this thesis are described above. For the remainder of the work Dr Benjamin Schulz contributed to the intellectual conception and design of the project, and critical revision of the written work.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

This dissertation would not have been possible without the guidance and help of several individuals who, in one way or another, contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, I would like to express my deepest gratitude to my principal supervisor, Dr Benjamin Schulz who has supported me throughout my PhD studies with his patience and knowledge whilst allowing me the room to work in my own way. He provided me with an outstanding atmosphere for doing research. Dr Ben has always been there through my ups and downs of experiment, giving me valuable advice and strong motivation. One simply could not wish for a better or friendlier supervisor.

Thanks to my secondary advisor, Associate Professor Bostjan Kobe and Dr Maja Bailey, for her academic guidance and exceptional moral support during my study and making this project a success. Without their encouragement and effort this thesis would not have been completed.

Thanks to my review panel members, Professor Ross Smith (Chair), Associate Professor Luke Guddat and Dr Amanda Nouwens for their thoughtful advice and support during different milestones of my PhD thesis. In my daily work, I have been blessed with a friendly and cheerful group of fellow students and staffs who always gave me more energy and strength to work in the lab and support for me in many ways; thanks to all past and present members of the Schulz group: Malcolm Lim, Anthony Stark, Margaret Davis, Ranjitha Naidu, Linnette Wong, Nikki Tan Yi Jie, Nur Hasanah Mohd Yusuf, Suresh Raman, Warren Jacobs, Amelia Tuffley, Siti Halimah Mahmud, Rokshareh Jowzaghi, Irene Yole Reto, Benjamin Yeo, Andri. All their support inspired me to hurdle all the obstacles in the completion of this research work. Special thanks to my best friend, Mohamed Fauzi Haroon from Australian Centre for Ecogenomics for his company, someone I could easily talk to and share my ideas during lunch break and for his valuable discussion not only pertaining to my project but other matters as well.

I would like to acknowledge the collaborative team on Crystallography research project including Associate Professor Bostjan Kobe and Dr Chiung Wen-Chang. The help, guidance and time they sacrifice have provided me have had a significant impact on my PhD.

Thanks to The University of Queensland International Postgraduate Research Scholarship and University of Queensland Research Scholarship that financially supported my tuition fee. Funding was also provided by National Health and Medical Research Council (NHMRC).

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation was simply impossible without them. I cannot find words to express my appreciation to my parents. I am indebted to my mother, Enon, and my dad, Jamaluddin who spared no effort throughout my life to provide the best possible environment for me to grow up and achieved my goals in both my academic and religious matters. I love u mom and dad. I admire the unconditional love and constant prayers and support of my wife Susanna and my son, Adam. She, like an intimate friend, shared all my joy and sorrow during my study and patiently putting up with my distraction while working on this project. It is unmentionable that how much support I got from my family here is Brisbane in coping with all the ups and downs and putting up with odd working hours and lost week-ends. She travels all the way from Wollongong to Brisbane just to be with me. I am grateful to my dear brothers, Faizal, Fadhli and Firdaus for their constant inspiration during my life. Being the second brother member of the family, they have always had a supportive and caring attitude towards me. None of my achievements would have been possible without their support.

Lastly, I would like to thank my mom in law, Sandra Knight for filling my life with joy, shelter and helping me quick to adapt to my new home Wollongong, and not to mention the internet usage I used at her home daily working on my thesis till night. Your support and encouragement was in the end what made this thesis possible.

Keywords

Glycoproteins, mass spectrometry, protein folding, protein translocation, yeast

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 030406 Proteins and Peptides, 45%

ANZSRC code: 060101 Analytical Biochemistry, 20%

ANZSRC code: 060112 Structural Biology, 35%

Fields of Research (FoR) Classification

For code: 0601, Biochemistry and Cell Biology, 50%

For code: 0304, Medicinal and Biomolecular Chemistry, 40%

For code: 0699, Other Biological Sciences, 10%

Table of Content

Abstract....ii

Declaration by author....v

Publications during candidature....vi

Publications included in this thesis....viii

Contributions by others to the thesis....xi

Statement of parts of the thesis submitted to qualify for the award of another degree....xi

Acknowledgements....xii

Keywords....xiv

Australian and New Zealand Standard Research Classifications (ANZSRC)....xiv

Fields of Research (FoR) Classification....xiv

Tables of Contents....xv

List of Figures....xix

List of Tables....xx

List of Abbreviations....xxi

Chapter 1: Introduction....1

1.1 Glycosylation....2

1.2 Asparagine (*N*)-linked glycosylation pathway....3

1.3 *N*-glycans in glycoprotein folding in the ER....6

1.4 Oligosaccharyltransferase (OTase)....7

1.4.1 OTase Subunits....7

1.4.1.1 Stt3....9

1.4.1.2 Ribophorin I....11

1.4.1.3 Ribophorin II....11

1.4.1.4 OST48....12

1.4.1.5 DAD1....13

1.4.1.6 OST4....13

1.4.1.7 MagT1/TUSC3.....14

1.5 Ost3p and Ost6p.....16

1.5.1 Ost3p and Ost6p define isoforms of OTase.....16

1.5.2 Molecular mechanisms of Ost3p and Ost6p function.....17

1.6 Research aims and significance.....21

1.7 References.....22

Chapter 2: Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation 28

2.1 Introduction.....29

2.2 Declaration on authorship.....29

2.3 Published peer-reviewed article.....29

"Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation"

Chapter 3: Selected reaction monitoring Mass Spectrometry for targeted peptide multi-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p.....70

3.1 Introduction.....71

3.2 Declaration on authorship.....71

3.3 Article to be submitted.....71

"Selected reaction monitoring Mass Spectrometry for targeted peptide mult-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p"

Chapter 4: Analysis of Congenital Disorder of Glycosylation-Id in a Yeast Model System shows diverse site-specific under-glycosylation of Glycoproteins.....87

4.1 Introduction.....88

4.2 Declaration on authorship.....88

4.3 Published peer-reviewed article.....88

"Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins"

Chapter 5: Sequence-based protein stabilization in the absence of glycosylation.....98

5.1 Introduction.....99

5.2 Declaration on authorship.....99

5.3 Published peer-reviewed article.....99

"Sequence-based protein stabilization in the absence of glycosylation"

Chapter 6: Transient protein-protein interaction between substrate peptide and yeast Oligosaccharyltransferase subunit Ost3/6p increase *N*-glycosylation.....116

6.1 Introduction.....117

6.2 Materials and Methods.....118

6.2.1 Recombinant DNA constructs.....118

6.2.2 Ligation-independent cloning.....119

6.2.3 Protein Expression and Purification.....120

6.2.3.1 Large-scale expression of His-tagged-Ost3L/ His-tagged-$Ost3_{N33K}$/ His-tagged-Ost3L-Linker-peptide/His-tagged-$Ost3_{N33K}$-Linker-peptide.....120

6.2.3.2 Purification of His_6-tagged of $Ost3_{N33K}$ and $Ost3_{N33K}$-Linker-peptide.....121

6.2.3.3 Purification of GST-tagged $Gas1_{85-97}$ peptide.....121

6.2.3.4 Purification of MBP-Ost3L.....122

6.2.4 Crystallization screening and optimization.....122

6.2.5 Kinetic analysis of ligand-analyte interaction.....123

6.3 Results

6.3.1 Rationale of construct design and cloning of pMCSG7-Ost3L/ pMCSG7 $Ost3_{N33K}$/ pMCSG7-Ost3L-Linker-Peptide/ pMCSG7-$Ost3_{N33K}$-Linker-Peptide.....124

6.3.2 Purification of $Ost3_{N33K}$ and $Ost3_{N33K}$-Linker-$Gas1_{85-97}$ peptide.....125

6.3.3 Purification of MBP-Ost3L.....127

6.3.4 Crystallization screening of MBP-Ost3L.....129

6.3.5 Purification of GST-$Gas1_{85-97}$ and GST pull-down of $Ost3_{N33K}$ protein.....131

6.3.6 Kinetic analysis of $Ost3_{N33K}$ versus GST-$Gas1_{85-97}$ peptide interaction.....132

6.4 Summary.....134

6.5 References.....136

Chapter 7: Conclusion and Future Directions.....138

7.1 Conclusion.....139

7.2 Future Directions.....142

7.3 References.....144

7.4 Appendix.....146

Lists of Figures

Figure 1.1: An overview of *N*-linked glycosylation pathway.....4

Figure 1.2: Structure of the mature *N*-glycan.....5

Figure 1.3: Yeast oligosaccharyltransferase complex.....8

Figure 1.4: Surface representation of the bacterial OTase PglB.....10

Figure 1.5: Peptide-binding groove of Ost6p and Ost3p.....18

Figure 1.6: Proposed model of Ost3/6p function in *N*-glycosylation.....20

Figure 6.1: Scheme expression of the pMCSG7 vector harbouring Ost3 or $Ost3_{N33K}$ with and without linker peptide.....124

Figure 6.2a: Purification of $Ost3_{N33K}$.....126

Figure 6.2b: Purification of $Ost3_{N33K}$-Linker-peptide.....127

Figure 6.3A: Gel filtration peak fractions.....128

Figure 6.3B: Purification of MBP-Ost3L.....128

Figure 6.4: Images of crystals of MBP-Ost3L.....130

Figure 6.5: Purification of $Ost3_{N33K}$ and GST-$Gas1_{85-97}$.....131

Figure 6.6: Kinetic analysis of GST-$Gas1_{85-97}$ peptide versus $Ost3_{N33K}$.....133

Figure 6.7: Kinetic analysis of $Ost3_{N33K}$ versus GST-$Gas1_{85-97}$.....133

Lists of Tables

Table 6.1: Primers used for Ost3L-/Ost6L-/Ost3$_{N33K}$/Ost3L Linker Peptide/Ost3$_{N33K}$ Linker Peptide.....119

Table 6.2: Primers used for pGEX2T-Gas1p$_{85-97}$.....119

List of Abbreviations

α	Alpha
ß	Beta
µ	micro
ºC	degrees Celsius
K_D	Dissociation constant
Aa	Amino acid
ANOVA	Analysis of variance
Asn	Asparagine
ALG	Asparagine-linked glycosylation
DAD1	Defender against apoptotic cell death
Dol-P	Dolichol-phosphate
Dol-PP-GlcNAc	Dolichyl-pyrophospate-*N*-acetylglucosamine
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
FPLC	Fast protein liquid chromatography
g	Gram(s)
GlcNAc	*N*-acetylglucosamine
Glc	Glucose
GRAVY	Grand average of hydropathy
GST	Glutathione S-transferase
His	Histidine
IPTG	Isopropyl-ß-D-Thio-Galactoside
LC-ESI-MS/MS	Liquid Chromatography Electrospray Ionization Tandem Mass Spectrometry
LIC	Ligation-independent cloning

LLO	Lipid linked oligosaccharide
M	molar (moles per litre)
m/z	Mass-to-charge ratio
MALDI-TOF	Matrix assisted laser desorption ionization time of flight
Man	Mannose
MBP	Maltose-Binding Protein
mg	milligram
mM	millimolar (millimoles per litre)
min	minute(s)
n	nano
Nm	Nanometer
NSMR	Non syndromic mental retardation
OTase	Oligosaccharyltransferase
PCR	Polymerase chain reaction
PglB	Protein glycosylation B
PMSF	Phenylmethylsulphonyl fluoride
PNgase F	Peptide-*N*-glycosidase
SDS-PAGE	Sodium docecyl sulphate polyacrylamide gel electrophoresis
Sec	Second
Ser	Serine
SRM-MS	Selected-reaction-monitoring mass spectrometry
STT3	Staurosporin and Temperature Sensitive 3
Thr	Threonine

Chapter 1

Introduction

1.1 Glycosylation

Protein glycosylation is the most common modification that regulates the structure and function of secretory and membrane proteins in eukaryotes (Herscovics and Orlean, 1993). This covalent addition of a carbohydrate molecule to a side chain of a polypeptide is an indispensable process in biology. Glycosylation aids in altering the properties of proteins, changing their stability, solubility and physical structure (Wormald and Dwek, 1999, Varki, 1993). In addition, the carbohydrate moieties function as recognition signals that stimulate protein targeting and influence cell-cell interactions and development of the organism (Herscovics and Orlean, 1993). Studies have shown that abnormalities in glycan expression are implicated as causative or incidental factors in relatively rare congenital diseases, widespread acquired diseases and cancer (Moremen et al., 2012, Ohtsubo and Marth, 2006, Schachter, 2000, Sell, 1990).

Glycosylation encompasses different types of glycans, including *N*-glycans, and *O*-glycans that are typically attached to cellular proteins (Ohtsubo and Marth, 2006). *N*-glycans are attached to asparagine (Asn) residues of proteins in sequons (Asn-X-Ser/Thr; X ≠ Pro), and help in protein folding in the endoplasmic reticulum (ER) (Schwarz and Aebi, 2011). *O*-glycans are linked to subset of serines and threonines in the Golgi (Schachter, 2000, Yan and Lennarz, 2005b). The final glycan structures on both *N*- and *O*-linked glycans are diversified during protein traffic through the Golgi. Many different glycosyltransferases carry out a similar basic reaction in which a sugar is transferred from an activated donor substrate to an acceptor molecule, usually another sugar residue that is part of an oligosaccharide under construction (Ohtsubo and Marth, 2006). These final glycan structures can affect protein function in diverse ways (Ohtsubo and Marth, 2006). This review focuses on the biosynthetic events of the *N*-glycosylation pathway as well as the function of oligosaccharyltransferase (OTase) subunits Ost3p and Ost6p in yeast *Saccharomyces cerevisiae*.

1.2 Asparagine (*N*)-linked glycosylation pathway

Asparagine (*N*)-linked glycosylation is a well-ordered process that involves the covalent attachment of oligosaccharides to selected target Asn residues (Helenius and Aebi, 2004). There are three main stages in the *N*-glycosylation mechanism: 1) the build up of a lipid-linked oligosaccharide (LLO) which is catalyzed by a series of glycosyltransferases encoding the asparagine-linked glycosylation (*ALG*) genes, 2) the transfer of the oligosaccharide to selected Asn residues of polypeptide chains and 3) the trimming, maturation and generation of glycan diversity in the Golgi (Figure 1.1) (Breitling and Aebi, 2013).

The biosynthesis of the LLO begins at the cytoplasmic surface of the ER membrane where dolichol-phosphate (Dol-P) lipid carrier acts as the glycosyl acceptor for *N*-acetylglucosamine (GlcNAc). The GlcNAc-phosphate is first added to Dol-P from nucleotide-activated UDP-GlcNac by the Alg7p glycosyltransferase encoded *ALG7*, to form the anhydride dolichyl-pyrophospate-GlcNAc (Dol-PP-GlcNAc) (Figure 1.1) (Breitling and Aebi, 2013). The presence of *ALG7* is essential in yeast and can be blocked by a drug called tunicamycin to prevent *N*-glycosylation pathway and cellular processes (Rine et al., 1983, Elbein, 1984, Kukuruzinska and Robbins, 1987, Breitling and Aebi, 2013). A second GlcNAc residue is then added to Dol-PP-GlcNAc by a protein complex encoded by *ALG13* and *ALG14*, to form Dol-PP-$GlcNAc_2$ (Bickel et al., 2005). Recent studies suggested that Alg7p, Alg13p and Alg14p assemble a complex that targets UDP-GlcNAc as the common substrate for efficient glycosylation by passing on LLO intermediates to consecutive active sites without diffusion (Noffz et al., 2009). Alg14p is thought to function as the key organizing subunit in this complex formation that interacts with Alg13p and, in addition, recruits Alg7p into the complex (Lu et al., 2012).

The following stages of LLO biosynthesis on the cytoplasmic surface involve the attachment of mannose (Man) residues from GDP-Man as a donor substrate to the Dol-PP-$GlcNAc_2$ (Figure 1.1) (Breitling and Aebi, 2013). The addition of the first Man to Dol-PP-$GlcNAc_2$ is catalyzed by the *ALG1* - encoded ß-1,4 mannosyltransferase to generate Dol-PP-$GlcNAc_2Man$ (Couto et al., 1984). Alg2p steps in next and transfers the two branching Man residues, first the α-1,3- and then the α-1,6-linked Man to the Dol-PP-$GlcNAc_2Man$ to form the lipid-linked $GlcNAc_2Man_3$ pentasaccharide (O'Reilly et al., 2006, Kämpf et al., 2009, Breitling and Aebi, 2013). The precise mechanism of the two different branching Man reactions performed by Alg2p is still poorly understood. The Dol-PP-$GlcNAc_2Man_3$ structure is recognized as a substrate for Alg11p, which extends the LLO by recruiting two α-1,2- linked Man residues to initiate the formation of lipid-linked $GlcNAc_2Man_5$ heptasaccharide (Cipollo et al., 2001, O'Reilly et al., 2006). The formation of Dol-PP-

$GlcNAc_2Man_5$ is the final product of *ALG* pathway on the cytoplasmic surface of the ER membrane (Figure 1.1) (Helenius and Aebi, 2004, Weerapana and Imperiali, 2006, Breitling and Aebi, 2013).

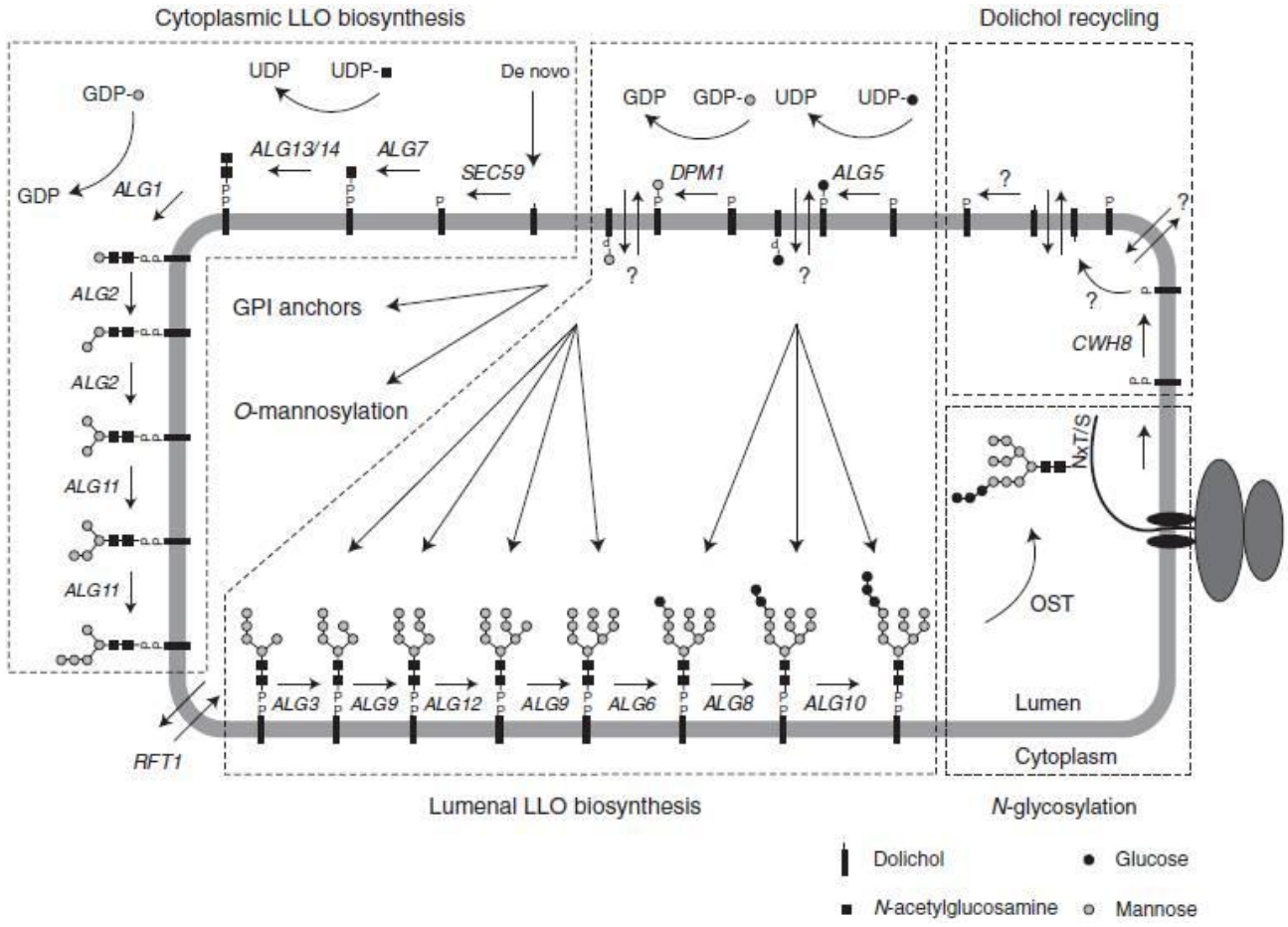


Figure 1.1: An overview of the *N*-linked glycosylation pathway is in the ER. Details are discussed in the text. The biosynthesis of lipid-linked oligosaccharide (LLO) is catalyzed by a series of glycosyltransferases encoded by asparagines-linked glycosylation (*ALG*) genes. The glycans begin to be built up at the cytoplasmic surface of the ER membrane by the addition of *N*-acetylglucosamine (GlcNAc)-phosphate to dolichol-phosphate (Dol-P) from nucleotide-activated UDP-GlcNAc. After the addition of a second GlcNAc residue to Dol-PP-GlcNAc, five mannoses (Man) are attached from GDP-Man to the LLO. After translocation into ER lumen, four additional Man and three glucose (Glc) residues are added to form a complete oligosaccharide ($Glc_3Man_9GlcNAc_2$). The completed oligosaccharide is then transferred en bloc to asparagine side chains of nascent polypeptides by the oligosaccharyltransferase (OTase) multiprotein complex (Breitling and Aebi, 2013).

Dol-PP-GlcNAc$_2$Man$_5$ is then translocated across the ER membrane by an ATP-independent bi-directional 'flippase', Rft1p (Figure 1.1) (Helenius et al., 2002). Then four additional Man and three glucose (Glc) residues are added, eventually yielding a full-length lipid-linked oligosaccharide Dol-PP-GlcNAc$_2$Man$_9$Glc$_3$ (Figure 1.2) (Helenius and Aebi, 2004, Breitling and Aebi, 2013). Unlike the glycosyltransferases acting on the cytoplasmic surface, the lumenal enzymes utilize the Dol-P-bound sugar substrates Dol-P-Glc and Dol-P-Man to elongate the LLO (Figure 1.1) (Breitling and Aebi, 2013). For example, Alg12p in α-1,6 linkage (Burda et al., 1999, Cipollo and Trimble, 2002) only initiates the c branch of the LLO when Alg9p has completed the b branch by capping one α-1,2-linked Man residue (Burda et al., 1999). Alg9p finally completes the c branch by the addition of α-1,2-linked Man residue (Burda et al., 1999, Frank and Aebi, 2005). Once the mannosylation of the LLO is completed, signalled by the presence of the complete b and c branches (Burda et al., 1999), Alg6p steps in to initiate the final glucosylation steps of the a branch (Reiss et al., 1996). Alg8p subsequently adds the second α-1,3-linked Glc residue to the LLO and the completion of the LLO is obtained by the attachment of α-1,2-linked Glc, catalysed by Alg10p (Burda and Aebi, 1998). This α-1,2-linked Glc residue represents an important final step in a linear assembly pathway of the branched oligosaccharide, ensuring the transfer of this completed oligosaccharide to proteins (Burda and Aebi, 1999). Absence of this terminal Glc residue impaires glycan recognition by OTase and correlates with a reduced *N*-glycosylation site occupancy in glycoprotein substrates (Breitling and Aebi, 2013). The result is an overall hypoglycosylation of the *N*-glycoproteome (Hülsmeier et al., 2007).

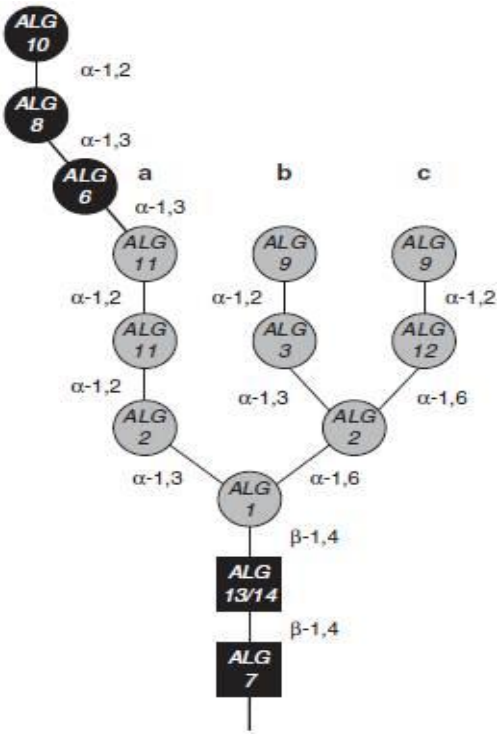


Figure 1.2: Structure of the mature *N*-glycan normally transferred to protein by OTase. The *N*-glycan structure consists of 14 carbohydrate residues. It contains two *N*-acetylglucosamines (GlcNAc, black squares), nine mannoses (Man, grey circles), and three glucoses (Glc, black circles). Linkage information is indicated between carbohydrate residues. The different residues are labelled with the genes encoding the glycosyltransferase that catalyzes the transfer of the respective residue. The structure of the glycan has three branches labelled a, b, and c (Breitling and Aebi, 2013).

The completed oligosaccharide $Glc_3Man_9GlcNAc_2$ (Figure 1.2) is subsequently transferred to the amide group of selected Asn residues with the consensus sequence Asn-Xaa-Ser/Thr (where Xaa is not Pro) of the polypeptide as it emerges into the ER lumen (Figure 1.1). This reaction is catalyzed by the oligosaccharyltransferase (OTase) (Kelleher and Gilmore, 2006, Mohorko et al., 2011, Schwarz and Aebi, 2011, Schulz, 2012).

1.3 *N*-glycans in glycoprotein folding in the ER

N-glycosylation is cotranslational; it occurs as the protein is being synthesized and it can affect protein folding (Helenius and Aebi, 2004). *N*-glycans assist in protein folding in the ER by their hydrophilic bulk lowering the entropy of the protein unfolded state and acting as nucleation sites for local folding of nascent glycopolypeptide chains (Helenius and Aebi, 2004, Schulz, 2012). *N*-glycans can also trigger a signalling pathway for particular proteins not folded correctly, and help to transport these to the ER resident thiol oxidoreductase ERp57 through the lectins calnexin and calreticulin (Oliver et al., 1999, Schulz, 2012).

Following the covalent linkage of the oligosaccharide from the dolichol-linked pyrophosphate donor to selected Asn side-chain, a series of processing reactions occurs. Firstly, the glycosidases (Glucosidases I and II) present in the lumen of the ER are responsible for the removal of all 3 glucose (Glc) sequentially (Helenius and Aebi, 2004). The presence or absence of Glc residues is crucial during glycoprotein folding as an indication that the glycoprotein is either properly folded and ready for release from the ER, or is not correctly folded and is to be reglucosylated by ER α-glucosyltransferase (Helenius and Aebi, 2004). The α-glucosyltransferase scans for improperly folded proteins in the ER and adds a Glc to their *N*-glycans. This allows glycoproteins to again be targeted to ERp57 to be refolded into proper conformation, or alternatively for glycoproteins to be eventually deglucosylated and degraded by the proteosome. The protein chaperones calnexin and calreticulin bind preferentially to $Glc_1Man_9GlcNAc_2$ and retain the glycoprotein in the ER for folding (Varki, 1993). Once protein folding is completed, glycoproteins traffic to the Golgi apparatus where further modifications in the form of trimming and addition of sugars occurs (Helenius and Aebi, 2004). The resulting *N*-linked oligosaccharides are diverse. Each has a common core region ($GlcNAc_2Man_3$) linked to Asn and originating from the dolichol-linked intermediate (Helenius and Aebi, 2004), while the terminal regions can be varied in different glycoproteins, cells, tissues and organisms.

1.4 Oligosaccharyltransferase (OTase)

In eukaryotic cells, OTase is an integral membrane protein that catalyzes *N*-linked glycosylation of nascent proteins in the lumen of the ER (Kelleher et al., 2003). OTase functions as the 'gatekeeper' to the secretory pathway, scanning the emerging nascent polypeptide for glycosylation 'sequons' (Asn-X-Thr/Ser; X≠Pro), and adding an oligosaccharyl moiety ($Glc_3Man_9GlcNAc_2$) (Figure 1.2) from the dolichol-linked pyrophosphate donor to selected Asn side-chain (Dempski and Imperiali, 2002, Yan and Lennarz, 2005a, Kelleher and Gilmore, 2006). Previous reports have demonstrated that glycosylation at Asn in the sequence Asn-X-Thr is approximately 40 times more efficient and frequently used than Asn-X-Ser sequons (Kasturi et al., 1995), and in a few cases, the glycosylation of Asn in Asn-X-Cys sequons has also been identified (Miletich and Broze, 1990, Zielinska et al., 2010). However, this is not an adequate predictor of glycosylation, as one-third of Asn in sequons in secreted proteins are not glycosylated (Schulz, 2012). It was also recently reported that *N*-glycosylation was observed on Asn-X-Val and Asn-Gly motifs (Zielinska et al., 2010). However, this result must be treated with extreme caution, given the propensity for non-catalyzed spontaneous deamidation (asparagines-aspartate conversion) is especially high at Asn-Gly sequences (Robinson et al., 2004, Palmisano et al., 2012).

While having a high degree of specificity with regard to the completed lipid-linked oligosaccharide donor (with structure $Glc_3Man_9GlcNAc_2$-P-P-dolichol) (Karaoglu et al., 2001), OTase accepts a large number of glycosylation sites situated on many different proteins. The large-scale proteomics study of the mouse *N*-glycoproteome allows detailed categorization of glycoproteins and revealed that glycosylation sites are preferentially located in surface-exposed loop and turn regions (Zielinska et al., 2010). In studies using SWISS-PROT protein sequence data, Apweiler et al. (1999) revealed that the majority of the sequon containing proteins in metazoans will be glycosylated and that more than half of all proteins in nature are glycoproteins. This indicates the structural diversity and sequence of proteins acceptor sites are phenomenally large (Apweiler et al., 1999).

1.4.1: OTase Subunits

In animals, plants and fungi, the OTase is a hetero-oligomeric complex. In yeast *Saccharomyces cerevisiae*, OTase is composed of eight different membrane–bound protein subunits that exist in three subcomplexes: Ost1p-Ost5p, Ost2p-Swp1p-Wbp1p, and Stt3p-Ost4p-Ost3p/Ost6p (Figure 1.3) (Lennarz, 2007, Yan and Lennarz, 2005b). The yeast OTase complex exists in two isoforms that differ with respect to either Ost3p or Ost6p (Schwarz et al., 2005, Spirig et al., 2005, Yan and

Lennarz, 2005a). Of these eight, five subunits (Ost1p, Ost2p, Stt3p, Wbp1p, and Swp1p) are encoded by essential genes and are important for viability (Heesen et al., 1992, Heesen et al., 1993, Zufferey et al., 1995, Silberstein et al., 1995b). Although the other genes (Ost3p, Ost4p, Ost5p, and Ost6p) are not essential for viability, deletion of any one causes an *N*-glycosylation defect (Chi et al., 1996, Karaoglu et al., 1995, Reiss et al., 1997, Knauer and Lehle, 1999). Thus, OTase requires all subunits for maximal enzyme activity (Harada et al., 2009).

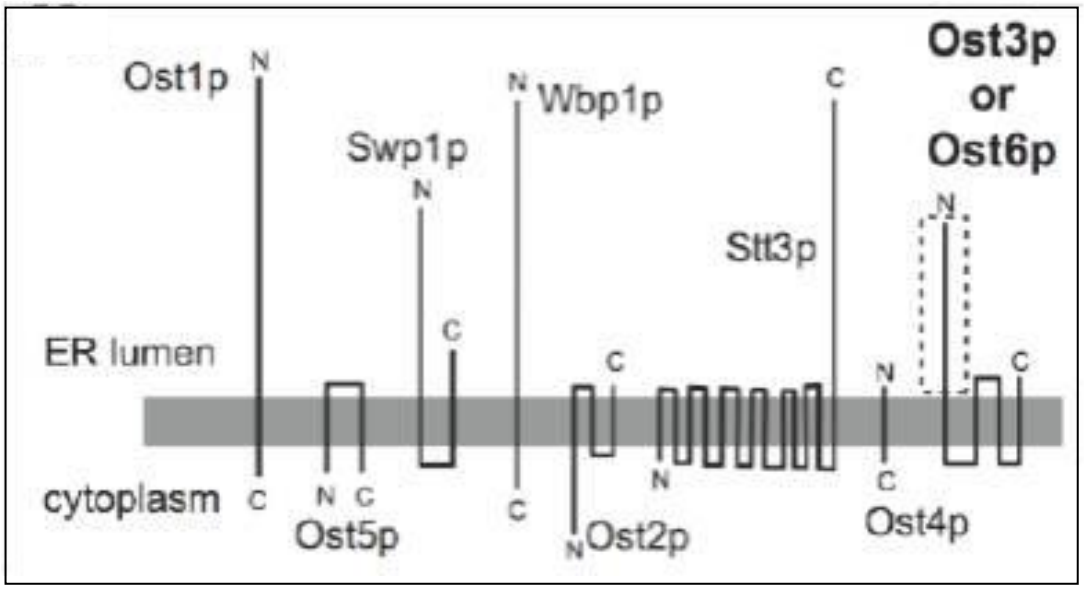


Figure 1.3: Yeast oligosaccharyltransferase (OTase) complex consists of eight different subunits. The scheme depicts the membrane topology of the individual subunits (amino (*N*) and carboxyl (*C*) termini) and the number of predicted transmembrane helices for each subunit. The dashed box indicates the site where thioredoxin-like folds are likely to occur in the ER lumenal domain of Ost3/6p, where only one of the respective subunits can be incorporated into a given OTase complex (Schulz and Aebi, 2009).

All of these subunits share high sequence identity and similarity with their human homologs (Kelleher et al., 1992, Kumar et al., 1994, Kumar et al., 1995, Breuer and Bause, 1995, Kelleher et al., 2003, Kelleher and Gilmore, 1997). Subunits of human OTase share a minimum of 90% pairwise sequence identity with the respective homolog in mouse, bovine, canine, rat and golden hamster OTase (Mohorko et al., 2011). They are ribophorin I (homolog of yeast Ost1p), ribophorin II (homolog of yeast Swp1p), OST48 (homolog of yeast Wbp1p), DAD1 (homolog of yeast Ost2p), STT3A/ STT3B (homolog of yeast Stt3p), MagT1/TUSC3 (homolog of yeast Ost3p/Ost6p) and OST4 (homolog of yeast Ost4p). In addition, all of these subunits have been characterized by genome-wide searches and have been demonstrated to be assembled together into a multimeric complex similar to the yeast OTase (Kelleher et al., 2003). The newly identified subunits KCP2 and

DC2 seem to associate with mammalian OST complexes (Shibatani et al., 2005, Wilson et al., 2011). However, there is no reported evidence about the functional roles of KCP2 and DC2 in the glycosylation process. In the following, we will describe the different subunits of the human OTase in more detail.

1.4.1.1: Stt3

Stt3 is the largest and most highly conserved transmembrane protein subunit of the eukaryotic OTase (Spirig et al., 1997, Mohorko et al., 2011). In lower eukaryotes, archaea and some bacteria, *N*-glycosylation is catalyzed by a single protein OTase, homologous with Stt3p (Castro et al., 2006, Izquierdo et al., 2009, Nasab et al., 2008). This reveals that Stt3p is the catalytic subunit because of its direct influence in the catalytic process (Burda and Aebi, 1999, Kelleher et al., 2003, Nilsson et al., 2003, Yan and Lennarz, 2005b). Although there is small or no similarity between the glycan structures transferred to the acceptor protein in eukaryotes, archaea and bacteria, the OTase catalytic domains are structurally and functionally related. The membrane topology of all Stt3 proteins consists of an *N*-terminal hydrophobic domain with thirteen transmembrane spans followed by a soluble *C*-terminal domain that is located in the lumen of the ER in eukaryotes or in the periplasm of gram-negative bacteria (Kelleher and Gilmore, 2006). The first X-Ray structure of a full length bacterial Stt3 homolog, the PglB from *Campylobacter lari*, revealed that the transmembrane domain is essential both for peptide binding and catalysis (Figure 1.4) (Lizak et al., 2011). This structure provides a molecular explanation for sequon recognition and suggestions for the catalytic mechanism in *N*-linked protein glycosylation (Lizak et al., 2011).

The structure of the PglB is anchored in the bacterial inner membrane via 13 transmembrane helical segments while a bulky globular domain faces into the periplasm (Lizak et al., 2011). Two cavities are found at the junction between the transmembrane and globular domains, and the acceptor peptide was found to be positioned across one of the cavities with its Asn side chain extending through a hole into the adjacent cavity, which is presumed to bind the LLO donor (Lizak et al., 2011). The nature of the interactions between the peptide acceptor and the active site cavity, the indirect catalytic role of a protein-bound divalent cation, and the conformational change in a loop region required for product release all provide novel insights into the mechanism of glycan transfer (Imperiali and Hendrickson, 1995). The additional subunits required for eukaryotic OTase activity presumably act to facilitate coupling to the Sec61 protein translocation pore and to bind recognized substrate polypeptide nearby glycosylation sequons in unfolded protein segments either during protein translocation into the ER or after translocation is completed (Chen and Helenius, 2000).

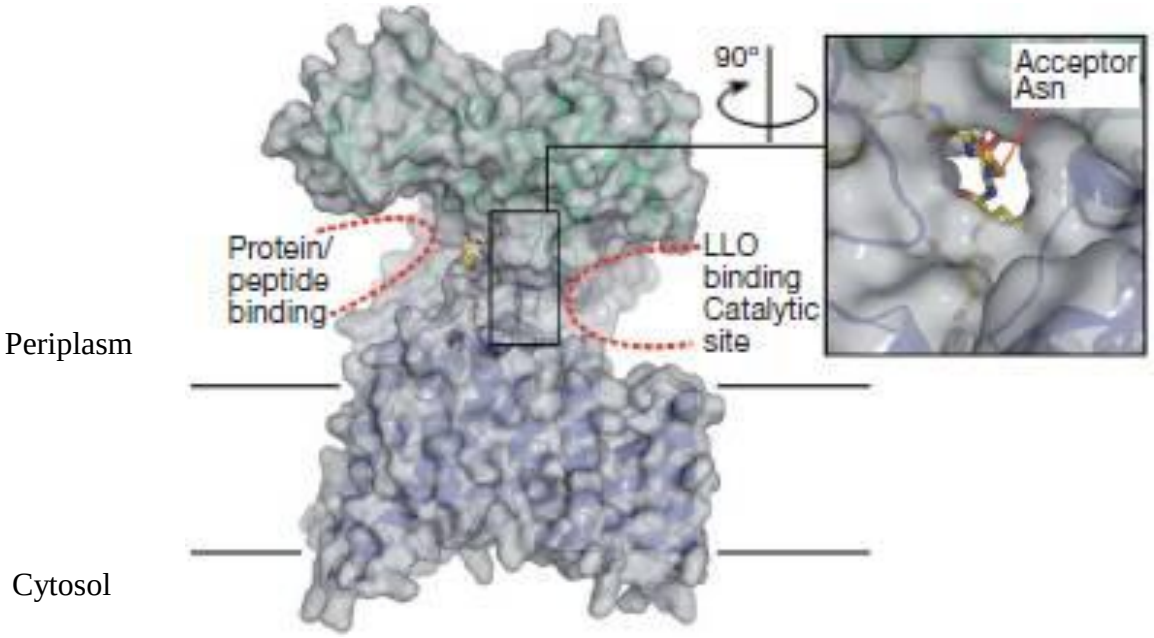


Figure 1.4: Surface representation of the bacterial OTase PglB. Cartoon in semi-transparent grey, with ribbons (transmembrane and periplasmic domains in blue and green respectively). There are two cavities at opposite sides of PglB indicated by dashed lines, providing access for substrates. The cavities are connected by a tunnel that harbours the acceptor asparagine (Lizak et al., 2011).

In humans, there are two isoforms of the catalytic subunit (Stt3A and Stt3B), and either Stt3A or Stt3B is incorporated into the human OTase complex (Kelleher et al., 2003, Mohorko et al., 2011). It has been established that Stt3A and Stt3B have different acceptor substrate preferences *in vivo* (Wilson and High, 2007). Wilson and High (2007) performed small interfering RNA knockdown experiments to investigate glycosylation sites of different membrane proteins in Stt3A- or Stt3B-deficient HeLa cells. The deletion of Stt3B results in hypoglycosylation of certain membrane proteins where as knocking down Stt3A causes complete loss of both the Stt3A and Stt3B isoforms and substantial reduction in *N*-glycosylation (Wilson and High, 2007). Interestingly, the differences between OTase harbouring either Stt3A or Stt3B were described more in a recent study (Ruiz-Canada et al., 2009) at the level of individual glycosylation sites in distinct substrates: Stt3B was observed to target sites glycosylated posttranslationally, and Stt3A catalyzed cotranslational glycosylation (Ruiz-Canada et al., 2009). This indicated that the actions of Stt3A or Stt3B differ with respect to substrate specificity as well as in the physical manner in the processing of polypeptide entering the ER lumen.

1.4.1.2: Ribophorin I

Ribophorin I is a highly abundant ER membrane protein (Marcantonio et al., 1984, Kreibich et al., 1978), together with OST48, were the first identified subunits of the mammalian OTase complex before the discovery of Stt3 (Kelleher et al., 1992). It was first thought that ribophorin I recruits a preassembled oligosaccharide from a membrane bound lipid donor to an acceptor Asn residue within a nascent polypeptide, but later studies clearly assigned the role to the Stt3 subunits (Yan and Lennarz, 2002, Nilsson et al., 2003). The membrane topology of ribophorin I consists of a single, predicted transmembrane helix; a large, lumenal domain; and a smaller, *C*-terminal cytosolic domain (Mohorko et al., 2011).

Cross-linking studies performed between ribophorin I and ribosomes led to insights into other possible functions of ribophorin I. Antibodies against the *C*-terminus of the ribophorin I cytosolic domain inhibit the ribosome from entering the ER membrane channel and thus block protein translocation (Yu et al., 1990). This indicated that ribophorin I may be involved in targeting the ribosome to the ER membrane (Yu et al., 1990). In addition, ribophorin I interacts with a subset of polypeptide substrates and presents them to the catalytic subunit Stt3 to facilitate efficient glycosylation of these proteins (Wilson and High, 2007, Wilson et al., 2008).

A protein sequence comparison revealed that human ribophorin I subunit is 28% identical to the yeast Ost1p (Silberstein et al., 1995b). The expression of the Ost1p is important for vegetative growth of haploid yeast, and mutant Ost1p demonstrated that Ost1p is required for *N*-linked glycoslation of proteins *in vivo* and for oligosaccharide transfer to acceptor peptides *in vitro* (Silberstein et al., 1995b).

1.4.1.3: Ribophorin II

Ribophorin II is an integral transmembrane glycoprotein of the rough ER that is highly conserved in mammals, and human ribophorin II shares >90% sequence similarity with ribophorin from mice, rat and pig (Crimaudo et al., 1987, Mohorko et al., 2011). The structure of ribophorin II is predicted to have a large *N*-terminal lumenal domain followed by a hydrophobic transmembrane domain with three α-helices and a small *C*-terminus region in the cytoplasm (Crimaudo et al., 1987). Both cytoplasmic and transmembrane domains of ribophorin II independently contain ER localization signals but, in order to mediate retention or strengthen ER localization, the transmembrane domain must be accompanied with fifteen amino acids at the lumenal region (Fu et al., 2000).

The expression of ribophorin II has also been implicated in drug resistant human breast cancer cells as a potential candidate predictive marker (Honma et al., 2008, Kurashige et al., 2012). Recently, Honma et al (2008) reported that downregulation of ripophorin II efficiently induced apoptosis in docetaxel-resistant human breast cancer cells in the presence of docetaxel, and silencing of ribophorin II inactivates glycosylation of the P-glycoprotein and decreased membrane localisation, thereby sensitising cancel cells to docetaxel (Honma et al., 2008). P-glycoprotein is indispensable in conferring drug resistance, and stable membrane bound P-glycoprotein requires complete glycosylation (Honma et al., 2008).

1.4.1.4: OST48

OST48, the third originally identified OTase subunit, was co-isolated from canine pancreas microsomal membranes (Kelleher et al., 1992). Canine OST48 and its homolog, Wbp1p, from the yeast *Saccharomyces cerevisiae* are found to be 25% identical in sequence (Silberstein et al., 1992). Like the ribophorins, OST48 is also highly conserved in mammals and human OST48 shares more than 90% sequence identity with other mammalian OST48 proteins (Yamagata et al., 1997) (Mohorko et al., 2011).

Human OST48 has 456 residues and the first 42 are predicted to have a signal sequence, followed by a large, ER lumenal domain. A single transmembrane helix constitutes OST48 as a type I membrane protein with a very small *C*-terminal cytosolic domain consisting of only nine residues (Mohorko et al., 2011). Interestingly, the *C*-terminal cytosolic domain of OST48 in canine was reported to interact with another OTase subunit, DAD1 (defender against apoptotic death) that may represent a fourth subunit of the mammalian OTase complex (Fu et al., 1997). Studies investigating these interactions among its subunits using yeast two-hybrid system have determined the interaction between lumenal domain of OST48 with ribophorin I and ribophorin II respectively but no direct interaction was observed between ribophorin I and ribophorin II (Fu et al., 1997). Knockdown of OST48 and DAD1 by small interfering RNA (siRNA) results in destabilization of both OTase complex isoforms (Roboti and High, 2012).

1.4.1.5: DAD1

DAD1 (defender against apoptotic cell death), a highly conserved OTase subunit, was initially discovered as a negative regulator of apoptosis in a temperature-sensitive mutant hamster cell line (tsBN7) (Nakashima et al., 1993). Rapid degradation of the DAD1 protein from a tsBN7 cells after shift to the restrictive temperature triggers apoptosis suggesting that DAD1 acts as cell death suppressor (Nakashima et al., 1993). The membrane topology of DAD1 consists of a small, 113-residue polypeptide with two predicted transmembrane spanning helices in the middle of the protein and both termini located in the cytoplasm (Mohorko et al., 2011). DAD1 protein is an authentic homolog of the yeast Ost2 protein and they share 40% sequence identity (Silberstein et al., 1995a). The extensive sequence homology between DAD1 and Ost2p and the *in vivo* and *in vitro* evidence demonstrating that Ost2p is an essential subunit of the yeast OTase complex suggested that DAD1 is a possible fourth subunit of the vertebrate complex (Silberstein and Gilmore, 1996, Sanjay et al., 1998).

DAD1 is a tightly associated subunit of the OTase both in the intact membrane and in the purified enzyme (Kelleher and Gilmore, 1997). It has been shown that the amino terminal domain of DAD1 interacts with *C*-terminus of OST48 (Fu et al., 1997). Interestingly, crosslinking experiments indicated that DAD1, OST48 and ribophorin II constitute a structural unit within the OTase core enzyme (Kelleher and Gilmore, 1997, Sanjay et al., 1998). They interact directly with each other (Fu et al., 1997, Kelleher and Gilmore, 1997), and it has been suggested that the resulting subcomplex assists in transferring the LLO donor to the catalytic Stt3 subunit (Beatson and Ponting, 2004, Kelleher et al., 2007). A mutation in the DAD1 subunit leads not only to the degradation of DAD1, but also destabilizes the other three subunits (Sanjay et al., 1998). This indicated that DAD1 is an essential component of the OTase complex and that its loss causes the rapid degradation of OST48 and also to a lesser extent of the ribophorins (Sanjay et al., 1998). The degradation of the OTase subunits eventually leads to complete loss of the *N*-glycosylation activity.

1.4.1.6: OST4

Human OST4 contains a very short lumenal domain with one transmembrane span followed by a cytoplasmic region of approximately ten residues. Mammalian OST4 has been tentatively identified in association with the purified canine OTase complex and it shows clear sequence identity to Ost4p, the smallest structural membrane protein of yeast OTase (Kelleher and Gilmore, 2006). In the yeast OTase complex, Ost4p regulates the incorporation of the two functionally equivalent, but mutually exclusive, subunits, Ost3p and Ost6p (Spirig et al., 2005). Point mutations in the *C*-

terminus of the transmembrane segment of Ost4p destabilize the interactions between Ost4p, Ost3p and Stt3p, causing severe growth defects in yeast (Kim et al. 2000, Kim et al., 2003).

Recently, Dumax-Vorzet et al (2013) provides an important role of OST4 that acts co-translational *N*-glycosylation by stabilising Stt3a-OTase isoforms. OST4 is assembled into native OTase complexes containing either the catalytic Stt3a or Stt3b isoforms. Knockdown of OST4 or Stt3A destabilizes the native OTase complexes and results in severe defect in *N*-glycosylation (Dumax-Vorzet et al., 2013).

1.4.1.7: MagT1/TUSC3

The human homologs of the yeast OTase subunits Ost3p and Ost6p are human MagT1 (IAP) and TUSC3 (N33) (MacGrogan et al., 1996, Kelleher et al., 2003). Human MagT1 and TUSC3 share 70% sequence identity and similar predicted membrane topology, containing an *N*-terminal lumenal domain followed by a transmembrane domain with four transmembrane helices (Kelleher et al., 2003, Mohorko et al., 2011). Based on sequence identity, Fetrow and co-workers proposed that MagT1 and TUSC3 also contain oxidoreductase activity (Fetrow et al., 2001) which correlates with later studies by Schulz (2009) on the lumenal domain of the yeast Ost6p. The important feature of the yeast Ost6p is a thioredoxin-like fold for the lumenal domain and revealed oxidoreductase activity for this domain which will be elaborated further in 1.5.2.

Recently, Mohorko (2014) solved the structures of N33/TUSC3-substrate complexes revealing the mode of substrate binding and that N33/TUSC3 may bind different substrate peptides in opposite orientations. Their structural and biochemical data show that N33/TUSC3 prefers peptides bearing a hydrophobic residue two residues away from the cysteine forming the mixed disulfide with N33/TUSC3 (Mohorko et al., 2014). Their results are parallel to our previously reported results showing that the thioredoxin-like domain of yeast Ost3p and Ost6 could form transient mixed disulfide bonds with cysteines in a model substrate glycoprotein, consistent with the function of N33/TUSC3 in modulating *N*-glycosylation for a subset of human glycoproteins by slowing glycoprotein folding (Mohd Yusuf et al., 2013).

Previous studies have reported that disruption of either Ost3p or Ost6p produces a minor defect in *N*-glycosylation (Karaoglu et al., 1995) and deletion of both OTase subunits results in severe underglycosylation of soluble/ membrane bound glycoproteins (Knauer and Lehle, 1999). Schulz and Aebi (2009) have also identified a number of specific glycosylation sites that are not efficiently glycosylated with lack of either Ost3p or Ost6p. Recently, Molinari et al. (2008) and Garshasbi et

al. (2008) reported mutations in the two OTase subunit genes encoding the human Ost3p/6p homologs (MagT1 and TUSC3) that can result in nonsyndromic mental retardation (NSMR). These data provide an association of NSMR with the congenital disorders of glycosylation syndromes which emphasize the crucial role of glycosylation activity in higher brain functions and cognitive development (Eklund and Freeze, 2006, Freeze, 2006). There were also other reports that revealed that TUSC3 inactivation leads to cancer development (Bova et al., 1993, Pils et al., 2006). In addition, the loss of functional MagT1 results in an inherited T-cell immunodeficiency (Li et al., 2011). At present, however, the precise role of glycosylation in NSMR and cancer is obscure and remains to be defined, but given the regulatory activities of yeast Ost3p and Ost6p, it is possible that human OTase MagT1 and TUSC3 also assist glycosylation of a subset of protein substrates required for correct human development and may be involved in cancer.

As many proteins require MagT1/TUSC3 for efficient glycosylation, the biology of affected cells and tissues would be altered in unpredictable and complex ways. MagT1 and TUSC3 have also been linked with magnesium transport. Previous studies have shown that expression of either TUSC3 or MagT1 can rescue growth in yeast lacking *ALR1*, encoding the major yeast magnesium transporter (Zhou and Clapham, 2009). Magnesium is a important metal ion that participates in many intracellular biochemical functions (Quamme, 2010).

Recent studies involving patients with a deficiency in MagT1 have shown that MagT1 is an important regulator of free intracellular magnesium, where magnesium functions as a second messenger in immune signalling activation of T-cell (Li et al., 2011). This indicates that MagT1 is required for T-cell activation, a finding which may help to open up potential treatment strategies for patients suffering from psoriasis, rheumatoid arthritis, or inflammatory bowel disease.

Together, this suggested that both human MagT1 and TUSC3 are associated with Mg^{2+} transport. However, TUSC3 has been detected physically associated with OTase, and lack of TUSC3 causes changes in protein glycosylation. While at least a fraction of MagT1 is present on the cell surface (Zhou and Clapham, 2009, Li et al., 2011), both TUSC3 and MagT1 share sequence characteristics with Ost3p and Ost6p that are important for their role in *N*-glycosylation. In the next section, we will explore the key characteristics and important roles of yeast Ost3p and Ost6p, which is our main focus in this project. This will focus on their different protein substrate binding specificities (Schulz and Aebi, 2009, Schulz et al., 2009, Jamaluddin et al., 2011) that will help us test our hypothetical model (Figure 1.6) of Ost3/Ost6p function in *N*-glycosylation.

1.5 Ost3p and Ost6p

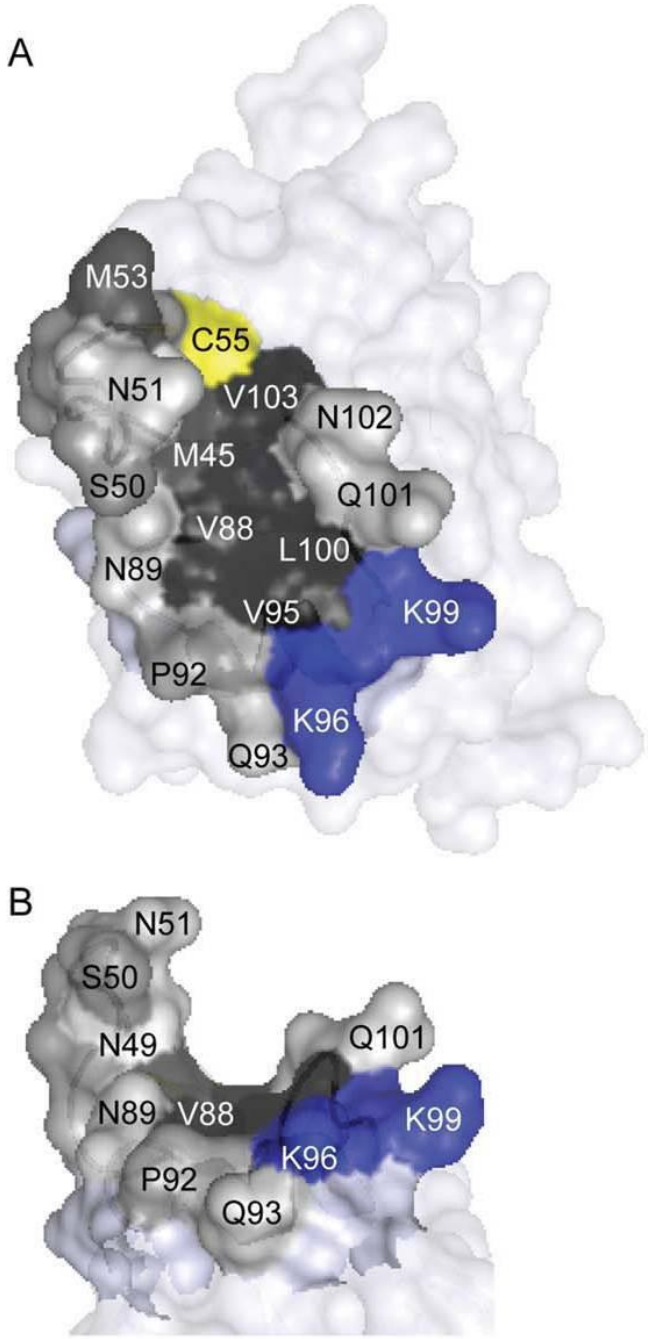
1.5.1: Ost3p and Ost6p define isoforms of OTase

The two best characterized accessory subunits of OTase are Ost3p and Ost6p. The sequence identity between Ost3p (UniProtKB number P48439; encodes the 34 kDa OTase subunit) and Ost6 (UniProtKB number Q03723; encodes the 37 kDa OTase subunit) is low, with 46% amino acid sequence homology and 21% sequence identity (Knauer and Lehle, 1999). However, based on the hydrophobicity plots of the paralogues Ost3p and Ost6p, both subunits contain similar *N*-terminal lumenal domains in the ER followed by four predicted transmembrane helices.

Two isoforms of OTase exist in yeast, determined by the presence of either Ost3p or Ost6p. The thioredoxin-like domain of Ost3p and Ost6p is likely to be located in the ER lumen (Figure 1.3) (Schulz and Aebi, 2009). Ost3p is thought to assemble into a more abundant form of the yeast OTase complex than Ost6p (Spirig et al., 2005). The function of Ost6p-complex was found to be crucial for cell wall integrity and temperature stress (Schwarz et al., 2005). Recently, genetic studies in the model plant *Arabidopsis* led to the discovery of an ER *N*-glycosylation pathway that requires the Ost3/6p subunits to allow *Arabidopsis* to tolerate abiotic stress, and also that the protein underglycosylation due to Ost3/6p deficiency results in the onset of ER stress (Farid et al., 2013). These findings indicated that changes in gene expression of subunits of the OTase complex and *N*-glycosylation activity are critical in response of an organism to a stressful environment. Ost3p and Ost6p can in part displace each other in the complex when overexpressed, suggesting a dynamic regulation of the complex formation (Schwarz et al., 2005).

1.5.2: Molecular mechanisms of Ost3p and Ost6 function

Previous studies predicted that the ER lumenal domains of Ost3p and Ost6p contain a thioredoxin-like fold and CxxC active-site motif (Fetrow et al., 2001). Thioredoxin-like folds are present in proteins with disulfide oxidoreductase activity and in the ER, generally catalyze disulfide bond formation and isomerisation during protein oxidative folding (Sevier and Kaiser, 2002). Recently, Schulz and Aebi (2009) established that the two OTase isoforms containing Ost3p or Ost6p differ in their protein substrate-specific activities at the level of individual glycosylation sites. Furthermore, in another experiment it was discovered that a specific peptide can transiently bind to the peptide-binding groove of the lumenal domain of Ost6p (Ost6L), but only when the CxxC active site of Ost6L is oxidized (Schulz et al., 2009). An interesting feature of the active-site loop of Ost6L is that it forms a peptide-binding groove when the CxxC is oxidized and becomes flexibly disordered when the CxxC is reduced (Schulz et al., 2009). The groove is a key candidate for a peptide-binding site, as it contains slightly basic and hydrophobic regions, an acidic patch, and encompasses several residues that could accommodate different substrate polypeptides (Figure 1.5) (Schulz et al., 2009). The structural and biochemical analyses of yeast Ost6L demonstrate the functional equivalence to the recently solved structures of N33/TUSC3-substrate complexes, in that the main function of these thioredoxin-like OST subunits is to increase glycosylation efficiency for a subset of glycosylation sequons in disulfide bonded glycoproteins (Mohd Yusuf et al., 2013, Mohorko et al., 2014). Interestingly, Mohorko and co-workers (2014) observed that the peptide-binding pocket of N33/TUSC3 shows no specific recognition of glycosylation sequons that could prevent the access of the catalytic OST subunit. This suggest a possible mechanism by yeast Ost3/Ost6 and N33/TUSC3 delays oxidative substrate folding through mixed disulfide formation and thereby increases the probability of sequon recognition and glycosylation by the catalytic subunit Stt3A/B as a nascent polypeptide chain passes the OTase complex during cotranslational translocation.



C

Ost6p $_{45}$**M**RGT**NSN**G**M**S**C**$_{55}$
Ost3p $_{41}$LFTATAPEIGC$_{51}$

Ost6p $_{86}$VD**VN**E----V**PQ**L**VK**DL**KLQNV**$_{103}$
Ost3p $_{89}$VNLEDPSKTIPKAF**Q**FF**Q**LNNV$_{110}$

Figure 1.5: Peptide-binding groove of *S. cerevisiae* Ost6p and Ost3p. Surface representation of **(A)** top and **(B)** side views of the ER lumenal domain of oxidized Ost6p (Ost6L; PDB code 3G7Y) with residues lining the peptide-binding groove coloured by hydropathy (black, hydrophobic to white hydrophilic); blue, basic; and yellow, cysteine. **(C)** Sequence alignment of sections of S. cerevisiae Ost6p and Ost3p, with surface-exposed residues in the Ost6p peptide binding groove and residues of Ost3p mutated in MBP-Ost3$_{Q103K,\ Q106K}$ variant bolded (Jamaluddin et al., 2011).

Our recent *in vitro* peptide binding assays have shown that Ost3p and Ost6p can bind stretches of polypeptide with complementary characteristics to their peptide- binding grooves (Jamaluddin et al., 2011). For example, the features of the peptides that interact with the groove of Ost6p are enriched with hydrophobic and acidic amino acid residues which perfectly complement the peptide-binding groove of Ost6p (hydrophobic base lined by basic residues). We also performed *in vitro* peptide binding assays using variant Ost3p and Ost6p proteins, which indeed indicated that the $Ost6_{K96Q,K99Q}$ variant abolished peptide binding while the $Ost3_{Q103K,Q106K}$ variant binds to one hydrophobic and acidic peptide that also bound to wild type Ost6p (Jamaluddin et al., 2011). This demonstrates that mutations made to the groove of Ost3p and Ost6p did affect peptide binding activities *in vitro*. To determine the physiological relevance of these *in vitro* interactions to *in vivo* *N*-glycosylation efficiency, it is necessary to investigate the effect of the amino acid residues in the peptide-binding grooves of Ost3p and Ost6p in OTase function *in vivo*.

The work in this thesis tests aspects of the hypothetical models of Ost3/6p function previously developed by Schulz et al. (2009). The central concept of the model in Figure 1.6 is that Ost3/6p function by binding nascent polypeptide, which slows protein folding and allows the catalytic site of OTase to access flexible nascent polypeptide substrate, which it can glycosylate efficiently (Schulz et al., 2009).

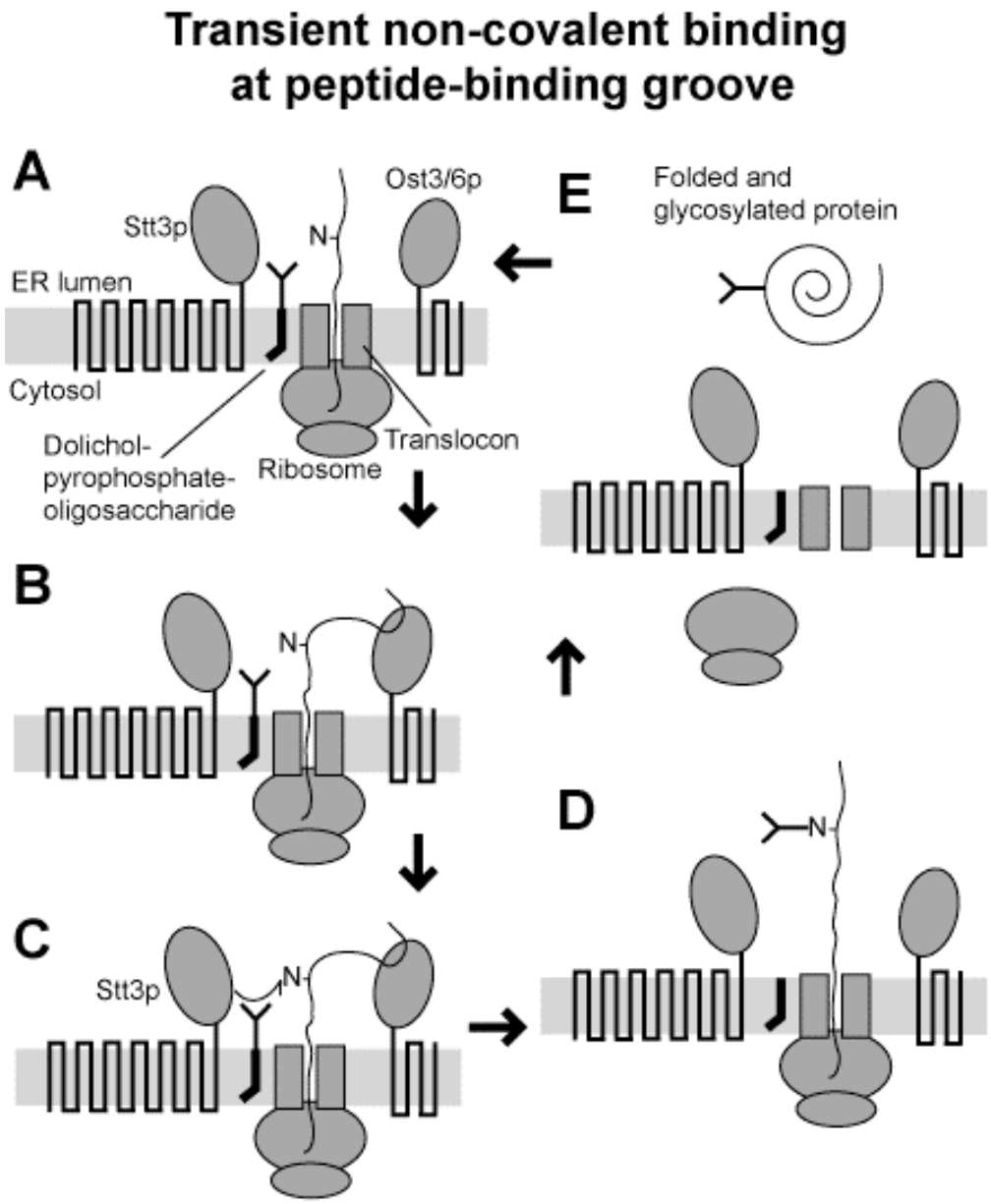


Figure 1.6: Proposed model of Ost3/6p function in *N*-glycosylation. (A) Nascent polypeptide chains enter through the translocon into the lumen of the ER. (B) Specific stretches of nascent polypeptides non-covalently bind at the peptide-binding groove. Translocation of the nascent polypeptide continues, with folding inhibited due to Ost3/6p binding constraints. (C) The catalytic site of OTase transfers glycan to the asparagine of a glycosylation sequon. (D) The binding is transient. Translocation and folding of the glycosylated polypeptide chain continues. (E) Glycosylated and folded protein (Schulz et al., 2009).

1.6 Research aims and significance

The specific aims of this PhD project are as follows:

1) Identifying the precise sites of interaction between the Gas1p substrate nascent polypeptide and the Ost3/6p peptide binding groove in *N*-glycosylation *in vivo* (Incorporated as Chapter 2)

2) To investigate the *in vitro* interactions between substrate polypeptide and Ost3p and Ost6p by peptide affinity chromatography and Selected Reaction Monitoring (SRM) Mass Spectrometry (Incorporated as Chapter 3)

3) Glycosite engineering to investigate the *in vitro* function of yeast Ost3p and Ost6 variants (Incorporated as Chapter 5)

4) Solving the crystal structure of Ost3p or Ost6p bound with a physiological ligand and measuring the strength of interaction between them (Incorporated as Chapter 6)

These specific aims will help us understand the molecular basis of the molecular mechanisms underlying glycosylation site selection by OTase activity, and details of the connection between glycoprotein folding and *N*-glycosylation in the endoplasmic reticulum.

1.7 References

APWEILER, R., HERMJAKOB, H. & SHARON, N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects,* 1473**,** 4-8.

BEATSON, S. & PONTING, C. P. 2004. GIFT domains: linking eukaryotic intraflagellar transport and glycosylation to bacterial gliding. *Trends Biochem Sci,* 29**,** 396-9.

BICKEL, T., LEHLE, L., SCHWARZ, M., AEBI, M. & JAKOB, C. A. 2005. Biosynthesis of lipid-linked oligosaccharides in Saccharomyces cerevisiae: Alg13p and Alg14p form a complex required for the formation of GlcNAc(2)-PP-dolichol. *J Biol Chem,* 280**,** 34500-6.

BOVA, G. S., CARTER, B. S., BUSSEMAKERS, M. J., EMI, M., FUJIWARA, Y., KYPRIANOU, N., JACOBS, S. C., ROBINSON, J. C., EPSTEIN, J. I., WALSH, P. C. & ET AL. 1993. Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. *Cancer Res,* 53**,** 3869-73.

BREITLING, J. & AEBI, M. 2013. N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol,* 5**,** a013359.

BREUER, W. & BAUSE, E. 1995. Oligosaccharyl Transferase Is a Constitutive Component of an Oligomeric Protein Complex from Pig-Liver Endoplasmic-Reticulum. *European Journal of Biochemistry,* 228**,** 689-696.

BURDA, P. & AEBI, M. 1998. The ALG10 locus of Saccharomyces cerevisiae encodes the α-1,2 glucosyltransferase of the endoplasmic reticulum: the terminal glucose of the lipid-linked oligosaccharide is required for efficient N-linked glycosylation. *Glycobiology,* 8**,** 455-462.

BURDA, P. & AEBI, M. 1999. The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta,* 1426**,** 239-57.

BURDA, P., JAKOB, C. A., BEINHAUER, J., HEGEMANN, J. H. & AEBI, M. 1999. Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology,* 9**,** 617-625.

CASTRO, O., MOVSICHOFF, F. & PARODI, A. J. 2006. Preferential transfer of the complete glycan is determined by the oligosaccharyltransferase complex and not by the catalytic subunit. *Proc Natl Acad Sci U S A,* 103**,** 14756-60.

CHEN, W. & HELENIUS, A. 2000. Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. *Mol Biol Cell,* 11**,** 765-72.

CHI, J. H., ROOS, J. & DEAN, N. 1996. The OST4 gene of Saccharomyces cerevisiae encodes an unusually small protein required for normal levels of oligosaccharyltransferase activity. *Journal of Biological Chemistry,* 271**,** 3132-3140.

CIPOLLO, J. F. & TRIMBLE, R. B. 2002. The Saccharomyces cerevisiae alg12Δ mutant reveals a role for the middle-arm α1,2Man- and upper-arm α1,2Manα1,6Man- residues of Glc3Man9GlcNAc2-PP-Dol in regulating glycoprotein glycan processing in the endoplasmic reticulum and Golgi apparatus. *Glycobiology,* 12**,** 749-762.

CIPOLLO, J. F., TRIMBLE, R. B., CHI, J. H., YAN, Q. & DEAN, N. 2001. The Yeast ALG11 Gene Specifies Addition of the Terminal α1,2-Man to the Man5GlcNAc2-PP-dolicholN-Glycosylation Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum. *Journal of Biological Chemistry,* 276**,** 21828-21840.

COUTO, J. R., HUFFAKER, T. C. & ROBBINS, P. W. 1984. Cloning and expression in Escherichia coli of a yeast mannosyltransferase from the asparagine-linked glycosylation pathway. *Journal of Biological Chemistry,* 259**,** 378-382.

CRIMAUDO, C., HORTSCH, M., GAUSEPOHL, H. & MEYER, D. I. 1987. Human ribophorins I and II: the primary structure and membrane topology of two highly conserved rough endoplasmic reticulum-specific glycoproteins. *EMBO J,* 6**,** 75-82.

DEMPSKI, R. E. & IMPERIALI, B. 2002. Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Current Opinion in Chemical Biology,* 6**,** 844-850.

DUMAX-VORZET, A., ROBOTI, P. & HIGH, S. 2013. OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation. *J Cell Sci,* 126**,** 2595-606.

EKLUND, E. A. & FREEZE, H. H. 2006. The congenital disorders of glycosylation: a multifaceted group of syndromes. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics,* 3**,** 254-63.

ELBEIN, A. D. 1984. Inhibitors of the biosynthesis and processing of N-linked oligosaccharides. *CRC Crit Rev Biochem,* 16**,** 21-49.

FARID, A., MALINOVSKY, F. G., VEIT, C., SCHOBERER, J., ZIPFEL, C. & STRASSER, R. 2013. Specialized roles of the conserved subunit OST3/6 of the oligosaccharyltransferase complex in innate immunity and tolerance to abiotic stresses. *Plant Physiol,* 162**,** 24-38.

FETROW, J. S., SIEW, N., DI GENNARO, J. A., MARTINEZ-YAMOUT, M., DYSON, H. J. & SKOLNICK, J. 2001. Genomic-scale comparison of sequence- and structure-based methods of function prediction: Does structure provide additional insight? *Protein Science,* 10**,** 1005-1014.

FRANK, C. G. & AEBI, M. 2005. ALG9 mannosyltransferase is involved in two different steps of lipid-linked oligosaccharide biosynthesis. *Glycobiology,* 15**,** 1156-1163.

FREEZE, H. H. 2006. Genetic defects in the human glycome. *Nature Reviews Genetics,* 7**,** 537-551.

FU, J., PIROZZI, G., SANJAY, A., LEVY, R., CHEN, Y., DE LEMOS-CHIARANDINI, C., SABATINI, D. & KREIBICH, G. 2000. Localization of ribophorin II to the endoplasmic reticulum involves both its transmembrane and cytoplasmic domains. *Eur J Cell Biol,* 79**,** 219-28.

FU, J., REN, M. & KREIBICH, G. 1997. Interactions among subunits of the oligosaccharyltransferase complex. *J Biol Chem,* 272**,** 29687-92.

HARADA, Y., LI, H., LI, H. & LENNARZ, W. J. 2009. Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proceedings of the National Academy of Sciences,* 106**,** 6945-6949.

HEESEN, S. T., JANETZKY, B., LEHLE, L. & AEBI, M. 1992. The Yeast Wbp1 Is Essential for Oligosaccharyl Transferase-Activity Invivo and Invitro. *EMBO J,* 11**,** 2071-2075.

HEESEN, S. T., KNAUER, R., LEHLE, L. & AEBI, M. 1993. Yeast Wbp1p and Swp1p Form a Protein Complex Essential for Oligosaccharyl Transferase-Activity. *EMBO J,* 12**,** 279-284.

HELENIUS, A. & AEBI, M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry,* 73**,** 1019-1049.

HELENIUS, J., NG, D. T., MAROLDA, C. L., WALTER, P., VALVANO, M. A. & AEBI, M. 2002. Translocation of lipid-linked oligosaccharides across the ER membrane requires Rft1 protein. *Nature,* 415**,** 447-50.

HERSCOVICS, A. & ORLEAN, P. 1993. Glycoprotein biosynthesis in yeast. *FASEB J,* 7**,** 540-50.

HONMA, K., IWAO-KOIZUMI, K., TAKESHITA, F., YAMAMOTO, Y., YOSHIDA, T., NISHIO, K., NAGAHARA, S., KATO, K. & OCHIYA, T. 2008. RPN2 gene confers docetaxel resistance in breast cancer. *Nat Med,* 14**,** 939-48.

HÜLSMEIER, A. J., PAESOLD-BURDA, P. & HENNET, T. 2007. N-Glycosylation Site Occupancy in Serum Glycoproteins Using Multiple Reaction Monitoring Liquid Chromatography-Mass Spectrometry. *Molecular & Cellular Proteomics,* 6**,** 2132-2138.

IMPERIALI, B. & HENDRICKSON, T. L. 1995. Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg Med Chem,* 3**,** 1565-78.

IZQUIERDO, L., SCHULZ, B. L., RODRIGUES, J. A., GUTHER, M. L. S., PROCTER, J. B., BARTON, G. J., AEBI, M. & FERGUSON, M. A. J. 2009. Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J,* 28**,** 2650-2661.

JAMALUDDIN, M. F. B., BAILEY, U.-M., TAN, N. Y. J., STARK, A. P. & SCHULZ, B. L. 2011. Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Science,* 20**,** 849-855.

KÄMPF, M., ABSMANNER, B., SCHWARZ, M. & LEHLE, L. 2009. Biochemical Characterization and Membrane Topology of Alg2 from Saccharomyces cerevisiae as a Bifunctional α1,3- and 1,6-Mannosyltransferase Involved in Lipid-linked Oligosaccharide Biosynthesis. *Journal of Biological Chemistry,* 284**,** 11900-11912.

KARAOGLU, D., KELLEHER, D. J. & GILMORE, R. 1995. Functional characterization of Ost3p. Loss of the 34-kD subunit of the Saccharomyces cerevisiae oligosaccharyltransferase results in biased underglycosylation of acceptor substrates. *J Cell Biol,* 130**,** 567-77.

KARAOGLU, D., KELLEHER, D. J. & GILMORE, R. 2001. Allosteric regulation provides a molecular mechanism for preferential utilization of the fully assembled dolichol-linked oligosaccharide by the yeast oligosaccharyltransferase. *Biochemistry,* 40**,** 12193-206.

KASTURI, L., ESHLEMAN, J. R., WUNNER, W. H. & SHAKIN-ESHLEMAN, S. H. 1995. The Hydroxy Amino Acid in an Asn-X-Ser/Thr Sequon Can Influence N-Linked Core Glycosylation Efficiency and the Level of Expression of a Cell Surface Glycoprotein. *Journal of Biological Chemistry,* 270**,** 14756-14761.

KELLEHER, D. J., BANERJEE, S., CURA, A. J., SAMUELSON, J. & GILMORE, R. 2007. Dolichol-linked oligosaccharide selection by the oligosaccharyltransferase in protist and fungal organisms. *J Cell Biol,* 177**,** 29-37.

KELLEHER, D. J. & GILMORE, R. 1997. DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase. *Proc Natl Acad Sci U S A,* 94**,** 4994-4999.

KELLEHER, D. J. & GILMORE, R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology,* 16**,** 47R-62R.

KELLEHER, D. J., KARAOGLU, D., MANDON, E. C. & GILMORE, R. 2003. Oligosaccharyltransferase Isoforms that Contain Different Catalytic STT3 Subunits Have Distinct Enzymatic Properties. *Molecular cell,* 12**,** 101-111.

KELLEHER, D. J., KREIBICH, G. & GILMORE, R. 1992. Oligosaccharyltransferase Activity Is Associated with a Protein Complex Composed of Ribophorin-I and Ribophorin-Ii and a 48kd Protein. *Cell,* 69**,** 55-65.

KNAUER, R. & LEHLE, L. 1999. The oligosaccharyltransferase complex from Saccharomyces cerevisiae. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. *J Biol Chem,* 274**,** 17249-56.

KREIBICH, G., ULRICH, B. L. & SABATINI, D. D. 1978. Proteins of rough microsomal membranes related to ribosome binding. I. Identification of ribophorins I and II, membrane proteins characteristics of rough microsomes. *J Cell Biol,* 77**,** 464-87.

KUKURUZINSKA, M. A. & ROBBINS, P. W. 1987. Protein glycosylation in yeast: transcript heterogeneity of the ALG7 gene. *Proc Natl Acad Sci U S A,* 84**,** 2145-9.

KUMAR, V., HEINEMANN, F. S. & OZOLS, J. 1994. Purification and characterization of avian oligosaccharyltransferase. Complete amino acid sequence of the 50-kDa subunit. *J Biol Chem,* 269**,** 13451-7.

KUMAR, V., KORZA, G., HEINEMANN, F. S. & OZOLS, J. 1995. Human oligosaccharyltransferase: isolation, characterization, and the complete amino acid sequence of 50-kDa subunit. *Arch Biochem Biophys,* 320**,** 217-23.

KURASHIGE, J., WATANABE, M., IWATSUKI, M., KINOSHITA, K., SAITO, S., NAGAI, Y., ISHIMOTO, T., BABA, Y., MIMORI, K. & BABA, H. 2012. RPN2 expression predicts response to docetaxel in oesophageal squamous cell carcinoma. *Br J Cancer,* 107**,** 1233-1238.

LENNARZ, W. J. 2007. Studies on oligosaccharyl transferase in yeast. *Acta Biochim Pol,* 54**,** 673-7.

LI, F.-Y., CHAIGNE-DELALANDE, B., KANELLOPOULOU, C., DAVIS, J. C., MATTHEWS, H. F., DOUEK, D. C., COHEN, J. I., UZEL, G., SU, H. C. & LENARDO, M. J. 2011. Second messenger role for Mg2+ revealed by human T-cell immunodeficiency. *Nature,* 475**,** 471-476.

LIZAK, C., GERBER, S., NUMAO, S., AEBI, M. & LOCHER, K. P. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature,* 474**,** 350-5.

LU, J., TAKAHASHI, T., OHOKA, A., NAKAJIMA, K.-I., HASHIMOTO, R., MIURA, N., TACHIKAWA, H. & GAO, X.-D. 2012. Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis. *Glycobiology,* 22**,** 504-516.

MACGROGAN, D., LEVY, A., BOVA, G. S., ISAACS, W. B. & BOOKSTEIN, R. 1996. Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22. *Genomics,* 35**,** 55-65.

MARCANTONIO, E. E., AMAR-COSTESEC, A. & KREIBICH, G. 1984. Segregation of the polypeptide translocation apparatus to regions of the endoplasmic reticulum containing ribophorins and ribosomes. II. Rat liver microsomal subfractions contain equimolar amounts of ribophorins and ribosomes. *J Cell Biol,* 99**,** 2254-9.

MILETICH, J. P. & BROZE, G. J. 1990. Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites. *Journal of Biological Chemistry,* 265**,** 11397-404.

MOHD YUSUF, S. N., BAILEY, U. M., TAN, N. Y., JAMALUDDIN, M. F. & SCHULZ, B. L. 2013. Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem Biophys Res Commun,* 432**,** 438-43.

MOHORKO, E., GLOCKSHUBER, R. & AEBI, M. 2011. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *Journal of Inherited Metabolic Disease,* 34**,** 869-878.

MOHORKO, E., OWEN, R. L., MALOJCIC, G., BROZZO, M. S., AEBI, M. & GLOCKSHUBER, R. 2014. Structural basis of substrate specificity of human oligosaccharyl transferase subunit N33/Tusc3 and its role in regulating protein N-glycosylation. *Structure,* 22**,** 590-601.

MOREMEN, K. W., TIEMEYER, M. & NAIRN, A. V. 2012. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol,* 13**,** 448-62.

NAKASHIMA, T., SEKIGUCHI, T., KURAOKA, A., FUKUSHIMA, K., SHIBATA, Y., KOMIYAMA, S. & NISHIMOTO, T. 1993. Molecular cloning of a human cDNA encoding a novel protein, DAD1, whose defect causes apoptotic cell death in hamster BHK21 cells. *Mol Cell Biol,* 13**,** 6367-74.

NASAB, F. P., SCHULZ, B. L., GAMARRO, F., PARODI, A. J. & AEBI, M. 2008. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol Biol Cell,* 19**,** 3758-68.

NILSSON, I., KELLEHER, D. J., MIAO, Y., SHAO, Y., KREIBICH, G., GILMORE, R., VON HEIJNE, G. & JOHNSON, A. E. 2003. Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J Cell Biol,* 161**,** 715-725.

NOFFZ, C., KEPPLER-ROSS, S. & DEAN, N. 2009. Hetero-oligomeric interactions between early glycosyltransferases of the dolichol cycle. *Glycobiology,* 19**,** 472-478.

O'REILLY, M. K., ZHANG, G. & IMPERIALI, B. 2006. In vitro evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in N-linked glycoprotein biosynthesis. *Biochemistry,* 45**,** 9593-603.

OHTSUBO, K. & MARTH, J. D. 2006. Glycosylation in cellular mechanisms of health and disease. *Cell,* 126**,** 855-67.

OLIVER, J. D., RODERICK, H. L., LLEWELLYN, D. H. & HIGH, S. 1999. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol Biol Cell,* 10**,** 2573-2582.

PALMISANO, G., MELO-BRAGA, M. N., ENGHOLM-KELLER, K., PARKER, B. L. & LARSEN, M. R. 2012. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res,* 11**,** 1949-57.

PILS, D., HORAK, P., KIRISITS, A., GLEISS, A., OMANN, M., PINTER, A., ZIELINSKI, C., MUSTEA, A., KOENSGEN, D., ZEILLINGER, R. & KRAINER, M. 2006. Hypermethylation of N33 (TUSC3) as independent molecular predictor for ovarian cancer. *Proceedings of the American Association for Cancer Research Annual Meeting,* 47**,** 760.

QUAMME, G. A. 2010. Molecular identification of ancient and modern mammalian magnesium transporters. *Am J Physiol Cell Physiol,* 298**,** C407-29.

REISS, G., HEESEN, S. T., ZIMMERMAN, J., ROBBINS, P. W. & AEBI, M. 1996. Isolation of the ALG6 locus of Saccharomyces cerevisiae required for glucosylation in the N-linked glycosylation pathway. *Glycobiology,* 6**,** 493-498.

REISS, G., TEHEESEN, S., GILMORE, R., ZUFFEREY, R. & AEBI, M. 1997. A specific screen for oligosaccharyltransferase mutations identifies the 9 kDa OST5 protein required for optimal activity in vivo and in vitro. *EMBO J,* 16**,** 1164-1172.

RINE, J., HANSEN, W., HARDEMAN, E. & DAVIS, R. W. 1983. Targeted selection of recombinant clones through gene dosage effects. *Proceedings of the National Academy of Sciences,* 80**,** 6750-6754.

ROBINSON, N. E., ROBINSON, Z. W., ROBINSON, B. R., ROBINSON, A. L., ROBINSON, J. A., ROBINSON, M. L. & ROBINSON, A. B. 2004. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *The journal of peptide research : official journal of the American Peptide Society,* 63**,** 426-36.

ROBOTI, P. & HIGH, S. 2012. The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation. *J Cell Sci,* 125**,** 3474-84.

RUIZ-CANADA, C., KELLEHER, D. J. & GILMORE, R. 2009. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell,* 136**,** 272-83.

SANJAY, A., FU, J. & KREIBICH, G. 1998. DAD1 Is Required for the Function and the Structural Integrity of the Oligosaccharyltransferase Complex. *Journal of Biological Chemistry,* 273**,** 26094-26099.

SCHACHTER, H. 2000. The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj J,* 17**,** 465-83.

SCHULZ, B. L. 2012. Beyond the sequon: sites of N-glycosylation. *inGlycosylation ed Petrescu S. 21-40Intech*.

SCHULZ, B. L. & AEBI, M. 2009. Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Molecular & Cellular Proteomics,* 8**,** 357-364.

SCHULZ, B. L., STIRNIMANN, C. U., GRIMSHAW, J. P. A., BROZZO, M. S., FRITSCH, F., MOHORKO, E., CAPITANI, G., GLOCKSHUBER, R., GRÜTTER, M. G. & AEBI, M. 2009. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proceedings of the National Academy of Sciences*.

SCHWARZ, F. & AEBI, M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol,* 21**,** 576-82.

SCHWARZ, M., KNAUER, R. & LEHLE, L. 2005. Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit. *FEBS Lett,* 579**,** 6564-6568.

SELL, S. 1990. Cancer-associated carbohydrates identified by monoclonal antibodies. *Hum Pathol,* 21**,** 1003-19.

SEVIER, C. S. & KAISER, C. A. 2002. Formation and transfer of disulphide bonds in living cells. *Nature Reviews Molecular Cell Biology,* 3**,** 836-847.

SHIBATANI, T., DAVID, L. L., MCCORMACK, A. L., FRUEH, K. & SKACH, W. R. 2005. Proteomic Analysis of Mammalian Oligosaccharyltransferase Reveals Multiple Subcomplexes that Contain Sec61, TRAP, and Two Potential New Subunits†. *Biochemistry,* 44**,** 5982-5992.

SILBERSTEIN, S., COLLINS, P. G., KELLEHER, D. J. & GILMORE, R. 1995a. The essential OST2 gene encodes the 16-kD subunit of the yeast oligosaccharyltransferase, a highly conserved protein expressed in diverse eukaryotic organisms. *J Cell Biol,* 131**,** 371-83.

SILBERSTEIN, S., COLLINS, P. G., KELLEHER, D. J., RAPIEJKO, P. J. & GILMORE, R. 1995b. The alpha subunit of the Saccharomyces cerevisiae oligosaccharyltransferase complex is essential for vegetative growth of yeast and is homologous to mammalian ribophorin I. *J Cell Biol,* 128**,** 525-36.

SILBERSTEIN, S. & GILMORE, R. 1996. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *The FASEB J,* 10**,** 849-58.

SILBERSTEIN, S., KELLEHER, D. J. & GILMORE, R. 1992. The 48-kDa subunit of the mammalian oligosaccharyltransferase complex is homologous to the essential yeast protein WBP1. *J Biol Chem,* 267**,** 23658-63.

SPIRIG, U., BODMER, D., WACKER, M., BURDA, P. & AEBI, M. 2005. The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology,* 15**,** 1396-1406.

SPIRIG, U., GLAVAS, M., BODMER, D., REISS, G., BURDA, P., LIPPUNER, V., TE HEESEN, S. & AEBI, M. 1997. The STT3 protein is a component of the yeast oligosaccharyltransferase complex. *Mol Gen Genet,* 256**,** 628-37.

VARKI, A. 1993. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology,* 3**,** 97-130.

WEERAPANA, E. & IMPERIALI, B. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology,* 16**,** 91R-101R.

WILSON, C. M. & HIGH, S. 2007. Ribophorin I acts as a substrate-specific facilitator of N-glycosylation. *J Cell Sci,* 120**,** 648-57.

WILSON, C. M., MAGNAUDEIX, A., YARDIN, C. & TERRO, F. 2011. DC2 and keratinocyte-associated protein 2 (KCP2), subunits of the oligosaccharyltransferase complex, are regulators of the gamma-secretase-directed processing of amyloid precursor protein (APP). *J Biol Chem,* 286**,** 31080-91.

WILSON, C. M., ROEBUCK, Q. & HIGH, S. 2008. Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proceedings of the National Academy of Sciences,* 105**,** 9534-9539.

WORMALD, M. R. & DWEK, R. A. 1999. Glycoproteins: glycan presentation and protein-fold stability. *Structure,* 7**,** R155-60.

YAMAGATA, T., TSURU, T., MOMOI, M. Y., SUWA, K., NOZAKI, Y., MUKASA, T., OHASHI, H., FUKUSHIMA, Y. & MOMOI, T. 1997. Genome organization of human 48-kDa oligosaccharyltransferase (DDOST). *Genomics,* 45**,** 535-40.

YAN, A. & LENNARZ, W. J. 2005a. Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology,* 15**,** 1407-1415.

YAN, A. & LENNARZ, W. J. 2005b. Unraveling the mechanism of protein N-glycosylation. *J Biol Chem,* 280**,** 3121-4.

YAN, Q. & LENNARZ, W. J. 2002. Studies on the Function of Oligosaccharyl Transferase Subunits: Stt3p IS DIRECTLY INVOLVED IN THE GLYCOSYLATION PROCESS. *Journal of Biological Chemistry,* 277**,** 47692-47700.

YU, Y. H., SABATINI, D. D. & KREIBICH, G. 1990. Antiribophorin antibodies inhibit the targeting to the ER membrane of ribosomes containing nascent secretory polypeptides. *J Cell Biol,* 111**,** 1335-42.

ZHOU, H. & CLAPHAM, D. E. 2009. Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development. *Proc Natl Acad Sci U S A,* 106**,** 15750-5.

ZIELINSKA, D. F., GNAD, F., WIŚNIEWSKI, J. R. & MANN, M. 2010. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell,* 141**,** 897-907.

ZUFFEREY, R., KNAUER, R., BURDA, P., STAGLJAR, I., TE HEESEN, S., LEHLE, L. & AEBI, M. 1995. STT3, a highly conserved protein required for yeast oligosaccharyl transferase activity in vivo. *EMBO J,* 14**,** 4949-60.

Chapter 2

Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation

2.1 Introduction to this publication

This chapter was published in *Molecular and Cellular Proteomics* as an original article. A model of Ost3/6p function has been proposed based on this *in vivo* and structural data in which we hypothesize that Ost3p and Ost6p transiently bind nascent polypeptide to increase the glycosylation efficiency of selected asparagines. In this paper, we tested a key prediction of this model using yeast genetics and glycoproteomics and established that this peptide-binding activity of Ost3/6p is physiologically relevant in *N*-glycosylation *in vivo*, suggesting that nascent polypeptide binds directly to the grooves in Ost3/6p. Further, we identified the precise sites of interaction between the Gas1p substrate nascent polypeptide and the Ost3/6p peptide binding groove *in vivo*, using a combination of site-directed mutagenesis and mass spectrometry glycoproteomics. Finally, using our *in vitro* peptide binding assay, we demonstrated and validated that a stretch of Gas1p we identify as interacting with Ost3/6p *in vivo* also interacted with Ost6p *in vitro*. Together, this study advances our understanding of Ost3/6p function in which they transiently bind stretches of nascent polypeptide substrate proximal and *N*-terminal to specific glycosylation sites, which inhibits local protein folding and thereby increases glycosylation efficiency.

2.2 Declaration on authorship

In accordance with The University of Queensland, the methods, results and discussion are presented in the form of a publication style in the peer-reviewed journal. This article was published in *Molecular and Cellular Proteomics*. The contributions of all authors are mentioned in Section 2.3

2.3 Published peer-reviewed article

"Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation"

Jamaluddin MF, Bailey UM, Schulz BL. Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation. *Mol Cell Proteomics*. **2014**, 13(12):3286-93.

Contributions:

This paper describes a model in which the Ost3p or Ost6p alternately bind to nascent proteins from the translocon thereby holding a "loop" of polypeptide in readiness for glycosylation. The Ost3p/Ost6p dichotomy enables the same OTase to preferentially glycosylate different sets of substrates, based on the partnering of peptides *N*-terminal to the glycosylation site with a binding groove in either Ost3p or Ost6p which show different binding characteristics.

This work presented here was done under the supervision of Dr Benjamin Schulz.

Dr Benjamin Schulz, Dr Maja Bailey and I planned experiments; Dr Maja Bailey and I performed *in vivo* yeast analyses; I performed *in vitro* peptide binding assays; Dr Benjamin Schulz, Dr Maja Bailey and I performed mass spectrometry; Dr Benjamin Schulz, Dr Maja Bailey and I wrote the manuscript. All authors read and approved the manuscript.

MCP Papers in Press. Published on August 12, 2014 as Manuscript M114.041178

Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance N-glycosylation.

M. Fairuz B. Jamaluddin[1], Ulla-Maja Bailey[1], and Benjamin L. Schulz[1,*]

[1] School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia

[*] To whom correspondence should be addressed: Benjamin L. Schulz, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia. Phone: +61 7 3365 4875 Fax: +61 7 3365 4273 Email: b.schulz@uq.edu.au

Running title: OTase subunits generate substrate competence

Abbreviations

OTase, oligosaccharyltransferase

Endo H, endoglycosidase H

GlcNAc, *N*-acetylglucosamine

Summary

Oligosaccharyltransferase is a multiprotein complex that catalyses asparagine-linked glycosylation of diverse proteins. Using yeast genetics and glycoproteomics we show that transient interactions between nascent polypeptide and Ost3p/Ost6p, homologous subunits of oligosaccharyltransferase, modulate glycosylation efficiency in a site-specific manner *in vivo*. These interactions are driven by hydrophobic and electrostatic complementarity between amino acids in the peptide-binding groove of Ost3p/Ost6p and the sequestered stretch of substrate polypeptide. Based on this dependence, we use *in vivo* scanning mutagenesis and *in vitro* biochemistry to map the precise interactions that affect site-specific glycosylation efficiency. We conclude that transient binding of substrate polypeptide by Ost3p/Ost6p increases glycosylation efficiency at asparagines proximal and C-terminal to sequestered sequences. We detail a novel mode of interaction between translocating nascent polypeptide and oligosaccharyltransferase in which binding to Ost3p/Ost6p segregates a short flexible loop of glycosylation-competent polypeptide substrate that is delivered to the oligosaccharyltransferase active site for efficient modification.

Introduction

Asparagine (N)-linked glycosylation is an essential post-translational modification of secretory and membrane proteins in eukaryota which also occurs in archaea and some bacteria (1). Oligosaccharyltransferase (OTase) is an integral membrane protein that catalyzes N-glycosylation of nascent polypeptides in the lumen of the endoplasmic reticulum (ER) (2). Eukaryotic OTase is physically associated with the translocon (3) and transfers oligosaccharide from a dolichol-pyrophosphate carrier onto asparagine side-chains of substrate polypeptides (2). The efficiency of glycosylation of asparagine residues is dramatically increased if they are present in glycosylation "sequons" (N-x-T/S; x≠P), the peptide recognition motif of the catalytic site of OTase (4). After transfer of glycan to protein the presence of N-glycosylation assists efficient glycoprotein folding in the ER intrinsically and by locally recruiting the disulfide isomerase ERp57 through the lectins calnexin and calreticulin (5). Correctly folded glycoproteins are free to traffic through the Golgi, where further modification and extension of glycan structures can occur (6). The precise glycan structures present on mature glycoproteins are often vital for their biological functions including in the immune response, embryonic development and cancer (6-8).

OTase in Bakers' yeast *Saccharomyces cerevisiae* is a multiprotein complex consisting of eight subunits (2). The catalytic site of OTase is located in the Stt3p subunit, while other protein subunits whose functions have been investigated are required for the integrity of the complex and for regulation of substrate specificity. It has been proposed that the requirement of the Ost1p subunit (human Ribophorin I) for efficient glycosylation of a subset of integral membrane

proteins (9) is due to direct physical association (10) to retain potential substrates in close proximity to Stt3p (11). The details and mechanisms of this association are not known. Ost3p and Ost6p are homologous proteins, incorporation of either of which into OTase defines two isoforms of the enzyme with distinct protein substrate specificities (Figure 1) (12-14). Efficient glycosylation of some asparagine residues requires the isoform of OTase that contains Ost3p (Ost3p-OTase), while glycosylation of other asparagines requires the isoform that contains Ost6p (Ost6p-OTase) (15). A model of Ost3p/Ost6p function in N-glycosylation site selection has been proposed (16), in which their ER-lumenal peptide-binding grooves transiently tether nascent polypeptide non-covalently or through mixed disulfides, inhibiting local protein folding and increasing the efficiency of glycosylation of nearby asparagine residues. Aspects of this model, including mixed-disulfide formation (17) and non-covalent peptide binding (18), have been tested *in vitro*. While it has been established that Ost3p-OTase and Ost6p-OTase have different polypeptide substrate preferences *in vivo* (12-15,19), the physiological relevance and details of any interactions between Ost3p/Ost6p and substrate polypeptide are unclear.

Here, we use *in vivo* yeast genetics, glycoproteomics and *in vitro* biochemistry to identify the precise sites of interaction between Ost3p/Ost6p and substrate polypeptides that impact the efficiency of N-glycosylation at diverse asparagines. Based on these data, we present a model in which transient binding of nascent polypeptide by Ost3p/Ost6p isolates newly translocating polypeptide from pre-translocated polypeptide, resulting in the formation of a short flexible loop of glycosylation-competent polypeptide substrate in close proximity to the active site of OTase for efficient modification.

Experimental Procedures

Cloning and mutagenesis. DNA encoding *OST3* or *OST6* including 1 kb up and downstream were amplified using PCR and cloned into the YEp352 vector using BamHI and SalI restriction enzymes, giving pOST3 and pOST6. DNA encoding *GAS1* including 1 kb up and downstream was amplified by PCR and cloned into the pRS415 vector using BamHI and SacI restriction enzymes, giving pGAS1. The thioredoxin-like ER lumenal domain of Ost6p and the mature soluble domain of Gas1p cloned as maltose binding protein (MBP) fusion proteins (pMAL-Ost6L and pMAL-Gas1, respectively) were used as described (18). Variants with amino acid replacements ($pOST3_{Q103K,Q106K}$, $pOST6_{K96Q,K99Q}$, $pGAS1_{K34Q}$, $pGAS1_{R47Q}$, $pGAS1_{R76Q}$, $pGAS1_{K82Q,K83Q}$, $pGAS1_{R90Q}$, $pGAS1_{K105Q}$, pMAL-$GAS1_{R90Q}$) were constructed as described (20).

Yeast strains. The *OST3* and *OST6* genes were genomically knocked-out in BY4742 (MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) using homologous recombination of transformed PCR products to produce strain Δ*ost3*/Δ*ost6* (MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0* Δ*ost3*::*KanMX* Δ*ost6*::*HisMX*). Yeast strain Δ*ost3*/Δ*ost6* was transformed with YEp352, pOST3, pOST6, $pOST3_{Q103K,Q106K}$ or $pOST6_{K96Q,K99Q}$ to produced yeast expressing only a single isoform of OTase (Fig. 1) (14). To determine the precise sites of interaction between Gas1p nascent polypeptide and the Ost6p peptide binding groove *in vivo*, pGAS1, $pGAS1_{K34Q}$, $pGAS1_{R47Q}$, $pGAS1_{R76Q}$, $pGAS1_{K82Q,K83Q}$, $pGAS1_{R90Q}$ or $pGAS1_{K105Q}$ were transformed into strain

Δ*ost3*/Δ*ost6* pOST6 or Δ*ost3*/Δ*ost6* pOST6$_{K96Q,K99Q}$. Yeast cells were grown to mid-log phase at 30°C in minimal medium (0.67% yeast nitrogen base and 2% glucose with appropriate supplements to allow growth).

Cell Wall Protein Sample Preparation. Analysis of site-specific glycosylation occupancy of cell wall glycoproteins was performed as described (15). Briefly, cells were mechanically lysed and insoluble cell wall material was reduced/alkylated and exhaustively washed. The resulting cell wall fraction is highly enriched in glycoproteins covalently linked to the cell wall polysaccharide through glycophosphatidylinositol remnants or alkaline sensitive linkages. N-glycans were released with endoglycosidase H (Endo H), leaving a single *N*-acetylglucosamine (GlcNAc) monosaccharide on previously glycosylated asparagines, and proteins were digested with trypsin. Peptides and glycopeptides were desalted with C18 ZipTips. Strains were analysed in biological triplicate from independent transformants. Peptides and deglycosylated peptides were detected with LC-ESI-MS/MS.

Peptide affinity chromatography. MBP-Ost6L, MBP-Gas1 and MBP-Gas1$_{R90Q}$ were expressed in *E.coli* carrying plasmids pMAL-OST6L, pMAL-GAS1 or pMAL-GAS1$_{R90Q}$, respectively, and purified by amylose-agarose affinity chromatography as described (18). MBP-Gas1 and MBP-Gas1$_{R90Q}$ were eluted in column buffer (20 mM Tris HCl pH 7.4, 200 mM NaCl, 1mM EDTA) with 1% SDS, while MBP-Ost6L was retained on the amylose-agarose resin. MBP-Gas1 and MBP-Gas1$_{R90Q}$ were reduced, alkylated, precipitated and digested with AspN as described (21). Residual AspN was inactivated by addition of EDTA to 1 mM. Peptide affinity chromatography

was performed as described (18). Briefly, peptides from equal amounts of MBP-Gas1 and MBP-$Gas1_{R90Q}$ in column buffer were combined and applied to amylose-agarose beads containing bound MBP-Ost6L, with the Ost6p CxxC motif either oxidized, or reduced by prior incubation with DTT. Peptides were washed from the affinity matrix by sequential gravity flow washing with column buffer and wash fractions collected. Peptides in the load and wash fractions were then detected by LC-ESI-MS/MS.

Mass Spectrometry and Data Analysis. Peptides were analysed by LC-ESI-MS/MS using a Prominence nanoLC system (Shimadzu) and TripleTof 5600 mass spectrometry with a Nanospray III interface (AB SCIEX) as described (21,22). Glycosylation occupancy was measured at previously identified glycosylation sites (15,21) using unmodified and GlcNAc-modified versions of the same peptide. This identification used ProteinPilot (AB SCIEX) searching the LudwigNR database with standard settings: Sample type, identification; Cysteine alkylation, acrylamide; Instrument, TripleTof 5600; Species, *S. cerevisiae* with common contaminants; ID focus, biological modifications; Enzyme, trypsin or AspN; Search effort, thorough ID. False discovery rate analysis using ProteinPilot was performed on all searches and peptides identified with greater than 99 % confidence and a local false discovery rate of less than 1 % were included for further analysis. The glycosylation occupancy at a given sequon was defined by the abundance as measured by peak intensity of the GlcNAc-modified peptide as a fraction of the sum of the GlcNAc-modified and unmodified versions of the same peptide (15). For *in vitro* peptide affinity chromatography analysis, the relative enrichment of each peptide in each wash fraction was determined as described (18) by the ratio of the abundance of each peptide. Interactions with the peptide-binding groove of Ost6p were determined for each peptide

by the log of the relative abundance of each peptide in each wash fraction between the oxidized and reduced versions of Ost6p. Experiments were performed in biological triplicate.

Results and Discussion

Ost3p and Ost6p peptide binding controls glycosylation

In vitro, Ost3p and Ost6p can transiently bind selected peptides with complementary characteristics to their peptide-binding grooves (16). The Ost6p groove is hydrophobic and lined by basic residues (K96 and K99), and binds peptides enriched in hydrophobic and acidic residues. Point mutations removing these basic residues in Ost6p ($Ost6_{K96Q,K99Q}$) abolish *in vitro* peptide binding, while addition of basic residues in place of the natively neutral residues at the equivalent positions in the Ost3p groove ($Ost3_{Q103K,Q106K}$) allows binding of a hydrophobic and acidic peptide that also binds to wild type Ost6p (18). This demonstrates that mutations to the groove of Ost3p and Ost6p affect peptide binding *in vitro*. To investigate the physiological relevance of these *in vitro* interactions to *in vivo* site-specific N-glycosylation efficiency, we constructed yeast strains expressing only a single isoform of OTase, wild type Ost3p-OTase, Ost6p-OTase, or variants of each with point mutations at their peptide-binding grooves ($Ost3_{Q103K,Q106K}$ or $Ost6_{K96Q,K99Q}$). We then used mass spectrometry to measure site-specific glycosylation occupancy of cell wall glycoproteins from these cells. As this method analyses mature glycoproteins in the yeast cell wall that have successfully folded and trafficked from the ER, it likely underestimates the affect of deleterious mutations to OTase (15,16). The analysis used Endo H to release N-glycans, leaving a single GlcNAc “tag” on previously glycosylated

asparagines. LC-ESI-MS/MS analysis could then detect GlcNAc-modified (glycosylated) and unmodified (non-glycosylated) forms of the same peptide to measure site-specific glycosylation occupancy at many different asparagines. For example, this approach detected both GlcNAc-modified and unmodified forms of peptide V91-K105 from Gas1p, which contains a potential glycosylation site at N95 (Fig. 2A,B).

We first compared yeast with only Ost3p-OTase or with only Ost6p-OTase. Consistent with previous reports (12,14,15), this analysis confirmed that certain asparagines required Ost3p-OTase for efficient glycosylation equivalent to wild type cells, while others required Ost6p-OTase. For example, glycosylation occupancy at N95 in Gas1p was significantly lower in yeast with only Ost6p-OTase compared to yeast with only Ost3p-OTase, indicating a requirement for Ost3p-OTase for efficient glycosylation at this site (Figure 2C,E). Equivalent analyses were performed in biological triplicate for the other robustly detected glycosylation sites in cell wall glycoproteins (Figure 3A column pOST3:pOST6, Table S1). This showed that efficient glycosylation at N253 in Gas1p required Ost6p-OTase, as has been previously reported (15). In contrast, Ost3p-OTase was required for efficient glycosylation at: N79 in Sag1p; N40, N57 and N95 in Gas1p; N177 in Crh1p, N304 and N328 in Ecm33p, N491 in Plb2p and N60 in Gas5p. While qualitatively consistent, our analysis here showed a stronger requirement for Ost3p-OTase than previous reports (15). Indeed, all asparagines that were glycosylated in wild type cells required the presence of either Ost3p or Ost6p for efficient glycosylation (Table S1), highlighting the importance of Ost3p/Ost6p-mediated glycosylation. Further, Gas1p contained three sites (N40, N57 and N95) that required Ost3p and one site (N253) that required Ost6p. As all of these sites are efficiently glycosylated in wild type yeast (with both Ost3p and Ost6p), this

result implies that both Ost3p-OTase and Ost6p-OTase isoforms are located concurrently at the same translocon. This is consistent with previous genetic and biochemical studies (13,15,16), although split-ubiquitin studies have proposed that Ost3p and Ost6p interact differentially with the Sec61 and Ssh1 translocons (23).

We next tested the effect of mutations in the peptide-binding grooves of Ost3p and Ost6p on *in vivo* site-specific glycosylation. For this, we compared site-specific glycosylation occupancy in yeast with only $Ost3_{Q103K,106K}$-OTase to yeast with only Ost3p-OTase; and in yeast with only $Ost6_{K96Q,K99Q}$-OTase to yeast with only Ost6p-OTase. These point mutations did affect *in vivo* glycosylation efficiency at selected sites. For example, glycosylation occupancy at N95 in Gas1p was higher in yeast with only $Ost6_{K96Q,K99Q}$-OTase compared to yeast with only Ost6p-OTase (Figure 2C,D), while occupancy at this site was lower in yeast with only $Ost3_{Q103K,Q106K}$-OTase compared to yeast with only Ost3p-OTase (Figure 2E,F). We performed equivalent analyses in biological triplicate for the other robustly detected glycosylation sites in cell wall glycoproteins (Figure 3A columns $pOST6_{QQ}$:pOST6 and $pOST3_{KK}$:pOST3, and Table S1). N57 and N95 in Gas1p and N328 in Ecm33p required Ost3p for efficient glycosylation and were under-glycosylated with Ost6p (Figure 3A,B,C,E and Table S1). However, this requirement for Ost3p was dependent on a neutral Ost3p/Ost6p peptide binding groove, as occupancy at Gas1p N57 and N95 was increased in the $Ost6_{K96Q,K99Q}$ variant and occupancy at all three sites was decreased in the $Ost3_{Q101K,Q103K}$ variant, relative to wild-type proteins (Figure 3A,B,C,E and Table S1). In contrast, N253 in Gas1p required Ost6p for efficient glycosylation, and occupancy at this site was increased in the $Ost3_{Q101K,Q103K}$ variant (Figure 3A,D and Table S1). These effects on site-specific glycosylation occupancy were consistent with growth of the corresponding yeast

strains. While the growth defect of *Δost3/Δost6* yeast was fully complemented by Ost3p and partially complemented by Ost6p, expression of $Ost3_{Q101K,Q103K}$ variant reduced growth relative to Ost3p (Figure S1), consistent with inefficient glycosylation at several asparagines in this strain (Figure 3). In summary, glycosylation by Ost6p-OTase at sites requiring Ost3p-OTase could be increased with point mutation to make Ost6p more like Ost3p, and vice versa. These results showed that point mutations that affect *in vitro* peptide binding by Ost3p and Ost6p also affected *in vivo* site-specific glycosylation occupancy, consistent with nascent polypeptide interacting directly with the grooves of Ost3p/Ost6p.

Mapping sites of interaction between substrate and Ost6p

Efficient glycosylation at N57 and N95 of Gas1p and N328 of Ecm33p required the absence of basic lysine residues in the peptide-binding groove of Ost3p/Ost6p (Figure 3A,B,C,E, Table S1). This suggested that specific basic amino acid residues in the Gas1p and Ecm33p nascent polypeptides caused electrostatic repulsion with Ost6p or $Ost3_{Q101K,Q103K}$, which had basic peptide-binding grooves (Figure 3F). We used this effect to identify the precise sites of interaction between the Gas1p substrate nascent polypeptide and the Ost6p peptide-binding groove *in vivo*.

We performed scanning mutagenesis of basic to neutral residues through the Gas1p sequence surrounding N57 and N95 and measured the occupancy of Gas1p glycosylation sites in each of these variant proteins in cells expressing Ost6p-OTase (see Figure S2). We predicted that mutagenesis of specific basic residues in Gas1p would eliminate the electrostatic repulsion and

allow binding to the basic peptide-binding groove of Ost6p, which would then increase the extent of glycosylation in a site-specific manner. Analysis of these cells showed a large and significant increase in occupancy at N95 in Gas1p only with mutation at R90 (Figure 4A,C and Table S2). This indicated that R90 in wild type Gas1p limited binding of this stretch of polypeptide to wild type Ost6p through electrostatic repulsion with the Ost6p peptide-binding groove, and that binding of Gas1p nascent polypeptide proximal to R90 with Ost6p was required for efficient glycosylation of N95. In contrast, occupancy at N57 was increased with mutation at R47, R76 or R90 (Figure 4A,B and Table S2). We interpreted this result as indicating that increased glycosylation of N57 could be achieved through binding at any one of these positions in Gas1p to Ost6p, or alternatively that some of these mutations destabilized Gas1p, leading to increased glycosylation independent of Ost6p binding. To distinguish between these possibilities we analysed the Gas1p variants in yeast with $Ost6_{K96Q,K99Q}$-OTase. Electrostatic repulsion between Ost6p and a section of Gas1p would require that each are positively charged, and point mutation to remove the charge on either one or on both would eliminate the repulsion. Therefore, increased occupancy with $Ost6_{K96Q,K99Q}$ relative to Ost6p for a particular Gas1p R variant would indicate that binding of that position in Gas1p to Ost6p was not required for glycosylation, while no change in occupancy would be consistent with that site in Gas1p binding to Ost6p to increase site-specific glycosylation occupancy. Analysis of these cells showed that in the $Gas1_{R90Q}$ variant, occupancy at N95 did not increase with $Ost6_{K96Q,K99Q}$ compared with Ost6p (Figure 4F and Table S3), consistent with the initial scanning mutagenesis (Figure 4A,C and Table S2) and with Gas1p nascent polypeptide proximal to R90 binding Ost6p to enhance glycosylation at N95. For N57 in Gas1p, occupancy increased with $Ost6_{K96Q,K99Q}$ relative to Ost6p for $Gas1_{R76Q}$ and $Gas1_{R90Q}$ but not for $Gas1_{R47Q}$ (Figure 4D,E and Table S3). This indicated that binding near R47

of Gas1p to Ost6p was required for efficient glycosylation of N57. Consistent with point mutations at R76 and R90 increasing glycosylation of N57 by destabilizing Gas1p, N57 is in a surface exposed loop of Gas1p in contrast to N95, which is relatively inaccessible (24) (Figure S3). Together, the data showed that transient binding to Ost3p/Ost6p of stretches of Gas1p nascent polypeptide proximal and N-terminal to N57 and N95 allowed efficient glycosylation at these sites (Figure 4G).

In vitro validation of in vivo peptide binding by Ost6p

To confirm that the increased glycosylation we observed in the Gas1p variants was due to increased physical association of stretches of Gas1p with Ost6p we performed *in vitro* peptide binding experiments. We expressed and purified Gas1p and $Gas1_{R90Q}$ as MBP-fusion proteins in *E. coli* and digested the proteins with AspN to obtain unglycosylated peptides similar to nascent polypeptide entering the ER lumen, which is the physiological substrate of OTase (Figure 5A). Peptides were applied to resin containing bound MBP-Ost6L, with the active-site CxxC motif either oxidized or reduced by pre-incubation with DTT. The peptide-binding groove is only present when the CxxC motif is oxidized, and peptides do not bind when this motif is reduced (16,18). We measured the MBP-Gas1 peptides present in each wash fraction by LC-ESI-MS/MS and observed increased retention of the robustly detected Gas1p77-108 peptide from the $Gas1_{R90Q}$ variant compared with wild type Gas1p with oxidized (Figure 5D) but not reduced (Figure 5C) Ost6p, compared to the load (Figure 5B). This indicated that $Gas1_{R90Q}$ mutation increased *in vitro* interaction with the Ost6p peptide-binding groove (Figure 5E), consistent with our *in vivo* results (Figure 3,4).

To gauge the generality of our results, we inspected the sequence N-terminal and proximal to glycosylation sites dependent on the presence of basic residues lining the peptide binding groove of either Ost3p or Ost6p (Figure 3A) and compared the characteristics of these sequences (Figure 6). Although Ost3p and Ost6p have different charge preferences, both require the presence of short stretches of hydrophobic amino acids in peptides for efficient binding (18,25). The closest hydrophobic sequence stretches N-terminal to N57 and N95 in Gas1p and N328 in Ecm33p were flanked by basic amino acids, consistent with efficient glycosylation of these sites requiring neutral peptide-binding grooves in Ost3p or Ost6p. In contrast, N253 in Gas1p had a hydrophobic stretch with acidic amino acids nearby, consistent with its preference for basic Ost3p or Ost6p.

Model of Ost3p/Ost6p function in vivo

Together, our data showed that Ost3p/Ost6p transiently bound hydrophobic stretches of nascent polypeptide substrate proximal and N-terminal to specific asparagines to increase their glycosylation efficiency. Based on these findings we propose a mechanistic model of Ost3p/Ost6p function in N-glycosylation (Figure 7). In this model, short hydrophobic stretches of nascent polypeptide exiting the translocon into the ER lumen (Figure 7A) bind to the peptide-binding groove of Ost3p/Ost6p independent of the location of nearby glycosylation sites (Figure 7B). This binding is non-covalent (18) or through a mixed disulfide (17) and depends on complementarity between the stretch of nascent polypeptide and the peptide-binding groove of Ost3p/Ost6p (Figures 4-6). Mutations at the thioredoxin-like CxxC active site motif of

Ost3p/Ost6p *in vivo* in yeast affect glycosylation at only ~10% of sites (16), suggesting the non-covalent mode of binding is dominant, and mixed disulfides form only when cysteines are appropriately positioned in stretches of bound nascent polypeptide. In either mode, this sequestration physically separates polypeptide that has already translocated into the ER lumen from newly translocating sequence. Continued translocation therefore produces a loop of nascent polypeptide with its ends constrained by the translocon exit and the Ost3p/Ost6p groove (Figure 7C). As OTase is physically associated with the translocon (3), and Ost3p/Ost6p is adjacent to Stt3p in the OTase complex (2,26), this configuration produces a flexible loop of nascent polypeptide in close proximity to the acceptor peptide binding site of Stt3p. This presents asparagines in glycosylation sequons to the active site of OTase in the form of short flexible peptides that are optimal substrates for efficient glycosylation (4,27,28) (Figure 7D,E). Ost3p/Ost6p peptide binding is transient (16-18) (Figure 7F), allowing continued translocation and folding of the glycosylated polypeptide (Figure 7G). Incorporation of Ost3p/Ost6p into OTase would therefore enhance glycosylation of diverse proteins, while their distinct peptide-binding specificities (17,18,25) would further increase the range of efficiently modified substrates.

Transient binding of short hydrophobic stretches of nascent polypeptide by Ost3p and Ost6p is reminiscent of the activity of co-translational Hsp70 chaperones such as cytoplasmic SSA/SSB (29) and ER lumenal Kar2p (BiP) (30), which assist co-translational folding (31,32). As binding to Ost3p or Ost6p is independent of the presence of a nearby glycosylation site, all stretches of nascent polypeptides with sufficient affinity to either Ost3p or Ost6p are likely transiently

sequestered. This may therefore facilitate proteome-wide co-translational folding of proteins translocated into the ER, independent of their glycosylation status.

References

1. Schwarz, F., and Aebi, M. (2011) Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struc Biol* **21**, 576-582
2. Kelleher, D. J., and Gilmore, R. (2006) An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R-62R
3. Harada, Y., Li, H., Li, H., and Lennarz, W. J. (2009) Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proc Natl Acad Sci U S A* **106**, 6945-6949
4. Lizak, C., Gerber, S., Numao, S., Aebi, M., and Locher, K. P. (2011) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355
5. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**, 1019-1049
6. Ohtsubo, K., and Marth, J. D. (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867
7. Marth, J. D., and Grewal, P. K. (2008) Mammalian glycosylation in immunity. *Nat Rev Immunol* **8**, 874-887
8. Schulz, B. L., Laroy, W., and Callewaert, N. (2007) Clinical laboratory testing in human medicine based on the detection of glycoconjugates. *Curr Mol Med* **7**, 397-416

9. Wilson, C. M., and High, S. (2007) Ribophorin I acts as a substrate-specific facilitator of N-glycosylation. *J Cell Sci* **120**, 648-657

10. Wilson, C. M., Kraft, C., Duggan, C., Ismail, N., Crawshaw, S. G., and High, S. (2005) Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon. *J Biol Chem* **280**, 4195-4206

11. Wilson, C. M., Roebuck, Q., and High, S. (2008) Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proceedings of the National Academy of Sciences* **105**, 9534-9539

12. Karaoglu, D., Kelleher, D. J., and Gilmore, R. (1995) Functional characterization of Ost3p. Loss of the 34-kD subunit of the Saccharomyces cerevisiae oligosaccharyltransferase results in biased underglycosylation of acceptor substrates. *J Cell Biol* **130**, 567-577

13. Schwarz, M., Knauer, M., and Lehle, L. (2005) Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit. *FEBS Lett* **579**, 6564-6568

14. Spirig, U., Bodmer, D., Wacker, M., Burda, P., and Aebi, M. (2005) The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **15**, 1396-1406

15. Schulz, B. L., and Aebi, M. (2009) Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Mol Cell Proteomics* **8**, 357-364

16. Schulz, B. L., Stirnimann, C. U., Grimshaw, J. P. A., Brozzo, M. S., Fritsch, F., Mohorko, E., Capitani, G., Glockshuber, R., Grütter, M. G., and Aebi, M. (2009)

Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc Natl Acad Sci U S A* **106**, 11061-11066

17. Mohd Yusuf, S. N., Bailey, U. M., Tan, N. Y. J., Jamaluddin, M. F. B., and Schulz, B. L. (2013) Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochemical and Biophysical Research Communications* **432**, 438-443

18. Jamaluddin, M. F. B., Bailey, U. M., Tan, N. Y. J., Stark, A. P., and Schulz, B. L. (2011) Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Sci* **20**, 849-855

19. Knauer, R., and Lehle, L. (1999) The oligosaccharyltransferase complex from Saccharomyces cerevisiae. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. *J Biol Chem* **274**, 17249-17256

20. Imai, Y., Matsushima, Y., Sugimura, T., and Terada, M. (1991) A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res* **19**, 2785

21. Bailey, U. M., Jamaluddin, M. F. B., and Schulz, B. L. (2012) Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J Proteome Res* **11**, 5376-5383

22. Bailey, U. M., Punyadeera, C., Cooper-White, J. J., and Schulz, B. L. (2012) Analysis of the extreme diversity of salivary alpha-amylase isoforms generated by physiological proteolysis using liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **911**, 21-26

23. Yan, A., and Lennarz, W. J. (2005) Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* **15**, 1407-1415

24. Hurtado-Guerrero, R., Schüttelkopf, A. W., Mouyna, I., Ibrahim, A. F., Shepherd, S., Fontaine, T., Latgè, J. P., and van Aalten, D. M. (2009) Molecular mechanisms of yeast cell wall glucan remodeling. *J Biol Chem* **284**, 8461-8469

25. Tan, N. Y., Bailey, U. M., Jamaluddin, M. F., Binte Mahmud, S. H., Raman, S. C., and Schulz, B. L. (2014) Sequence-based protein stabilization in the absence of glycosylation. *Nat Comms* **5**

26. Yan, Q., and Lennarz, W. J. (2002) Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J Biol Chem* **277**, 47692-47700

27. Gerber, S., Lizak, C., Michaud, G., Bucher, M., Darbre, T., Aebi, M., Reymond, J. L., and Locher, K. P. (2013) Mechanism of bacterial oligosaccharyltransferase: in vitro quantification of sequon binding and catalysis. *J Biol Chem* **288**, 8849-8861

28. Holst, B., Bruun, A. W., Kielland-Brandt, M. C., and Winther, J. R. (1996) Competition between folding and glycosylation in the endoplasmic reticulum. *EMBO J* **15**, 3538-3546

29. Willmund, F., del Alamo, M., Pechmann, S., Chen, T., Albanèse, V., Dammer, E. B., Peng, J., and Frydman, J. (2013) The Cotranslational Function of Ribosome-Associated Hsp70 in Eukaryotic Protein Homeostasis. *Cell* **152**, 196-209

30. Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991) Peptide-binding specificity of the molecular chaperone BiP. *Nature* **353**, 726-730

31. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., and Hartl, F. U. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **82**, 323-355

32. Kampinga, H. H., and Craig, E. A. (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* **11**, 579-592

Footnotes

This work was supported in part by a project grant to B.L.S. from the National Health and Medical Research Council (631615). B.L.S was supported in part by a Career Development Fellowship from the National Health and Medical Research Council (APP1031542).

Figure Legends

Figure 1. Overview of experimental manipulation of yeast OTase isoforms. (**A**) OTase in wild type yeast has eight subunits, with two isoforms defined by incorporation of either of the homologous Ost3p or Ost6p subunits. Yeast with only (**B**) Ost3p-OTase or (**C**) Ost6p-OTase were constructed by genomic deletion of *OST3* and *OST6* with overexpression of either Ost3p or Ost6p. (**D**) Yeast with a single variant OTase isoform were constructed by overexpression of variant Ost3p or Ost6p, for example $Ost3_{Q103K,Q106K}$.

Figure 2. LC-ESI-MS/MS identification and relative quantification of glycosylation occupancy at N95 in Gas1p. ProteinPilot matched fragmentation of (**A**) GlcNAc-modified and (**B**) unmodified peptide V91-K105 from Gas1p. Single ion chromatograms corresponding to GlcNAc-modified (GlcNAc-N, black) and unmodified (N, gray) peptide V91-K105 from Gas1p ($[M+3H]^{3+}$ ions at an *m/z* of 666.98 and 559.30, respectively) from yeast with only (**C**) Ost6p-OTase (Δ*ost3*/Δo*st6* pOST6), (**D**) $Ost6_{K96Q,K99Q}$-OTase (Δ*ost3*/Δ*ost6* $pOST6_{K96Q,K99Q}$), (**E**) Ost3p-OTase (Δ*ost3*/Δ*ost6* pOST3), or (**F**) $Ost3_{Q103K,Q106K}$-OTase (Δ*ost3*/Δ*ost6* $pOST3_{Q103K,Q106K}$), Horizontal gray lines are aligned with the peak height of the unmodified peptide in (C) between (C) and (D), and in (F) between (E) and (F).

Figure 3. Ost3p and Ost6p peptide-binding groove variants affect *in vivo* site-specific glycosylation. (**A**) Relative occupancy at different glycosylation sites (e.g. N79 in Sag1p) in

Δ*ost3*/Δ*ost6* cells with: pOST3 relative to pOST6 (pOST3:pOST6); $pOST6_{K96Q,K99Q}$ relative to pOST6 ($pOST6_{QQ}$:pOST6); and $pOST3_{Q103K,Q106K}$ relative to pOST3 ($pOST3_{KK}$:pOST3). Ratio is mapped from yellow (increase) to blue (decrease) or black (no change), see color key inset. *, $P<0.05$ ANOVA. Occupancy of (**B**) N57 in Gas1p (**C**) N95 in Gas1p, (**D**) N253 in Gas1p and (**E**) N328 in Ecm33p in various yeast strains. Values are mean. Error bars are s.e.m.. *, $P<0.05$ ANOVA. (**F**) Cartoon surface representation of proposed binding of hydrophobic and neutral (black) but not hydrophobic and basic (blue) stretches of nascent polypeptide by the basic peptide binding groove of Ost6p (PDB ID 3G7Y (16,18)).

Figure 4. Mapping sites of interaction between Ost6p and Gas1p. (**A**) Relative occupancy at N40, N57, N95 and N253 in Gas1p variants compared to wild type Gas1p expressed in Δ*ost3*/Δ*ost6* pOST6 cells. Colour as in Fig. 3A. *, $P<0.05$ ANOVA. Occupancy of (**B**) N57 in Gas1p and (**C**) N95 in Gas1p in various yeast strains. Values are mean. Error bars are s.e.m..*, $P<0.05$ ANOVA. (**D**) Relative occupancy at N57 and N95 in Gas1p variants expressed in Δ*ost3*/Δ*ost6* $pOST6_{K96Q,K99Q}$ compared to Δ*ost3*/Δ*ost6* pOST6. Colour as in Fig. 3A. Occupancy of (**E**) N57 in Gas1p and (**F**) N95 in Gas1p in various Gas1p variants expressed in Δ*ost3*/Δ*ost6* pOST6 (black) or Δ*ost3*/Δ*ost6* $pOST6_{K96Q,K99Q}$ (white). Values are mean. Error bars are s.e.m.. *, $P<0.05$ ANOVA. (**G**) Position of basic-neutral point mutations (blue) in Gas1p that increased occupancy at specific glycosylation sites (green).

Figure 5. Non-covalent peptide binding by Ost6p *in vitro*. (**A**) Aligned AspN peptides from MBP-Gas1 and MBP-$Gas1_{R90Q}$ detected by LC-ESI-MS/MS. Extracted ion chromatograms of

peptide from MBP-Gas1 (solid) and MBP-Gas1$_{R90Q}$ (dashed) in (**B**) load, and second wash fraction with (**C**) reduced MBP-Ost6 and (**D**) oxidized MBP-Ost6. (**E**) Relative enrichment of the Gas1$_{77\text{-}108}$ versus the Gas1$_{R90Q\ 77\text{-}108}$ peptide by columns with attached oxidized MBP-Ost6. Data are average of triplicates. Error bars indicate range.

Figure 6. Conservation of sequence characteristics at sites of Ost3p/Ost6p interaction. Validated and predicted stretches of substrate proteins Gas1p and Ecm33p that bind to Ost3p or Ost6p. Underlined, glycosylation sequon. Grey shading, extended hydrophobic sequence. Blue, basic amino acid. Red, acidic amino acid.

Figure 7. Mechanistic model of Ost3p/Ost6p function in N-glycosylation. (**A**) Partially translocated and folded nascent polypeptide enters the ER lumen from the translocon. (**B**) A stretch of nascent polypeptide (red) with affinity for Ost3p/Ost6p is sequestered at the peptide-binding groove. (**C**) Continued translocation produces a flexible loop of nascent polypeptide kept physically separate from previously translocated polypeptide. (**D**) Glycosylation sequons in the loop are presented to the acceptor-binding site of Stt3p in a glycosylation-competent form and (**E**) efficiently glycosylated. (**F**) Non-covalent or mixed-disulfide binding of nascent polypeptide to Ost3p/Ost6p is transient, and (**G**) translocation and folding continues. Ost3p/Ost6p and Stt3p are physically adjacent in OTase (2) but are shown separately for clarity. See text for details.

Figure 1

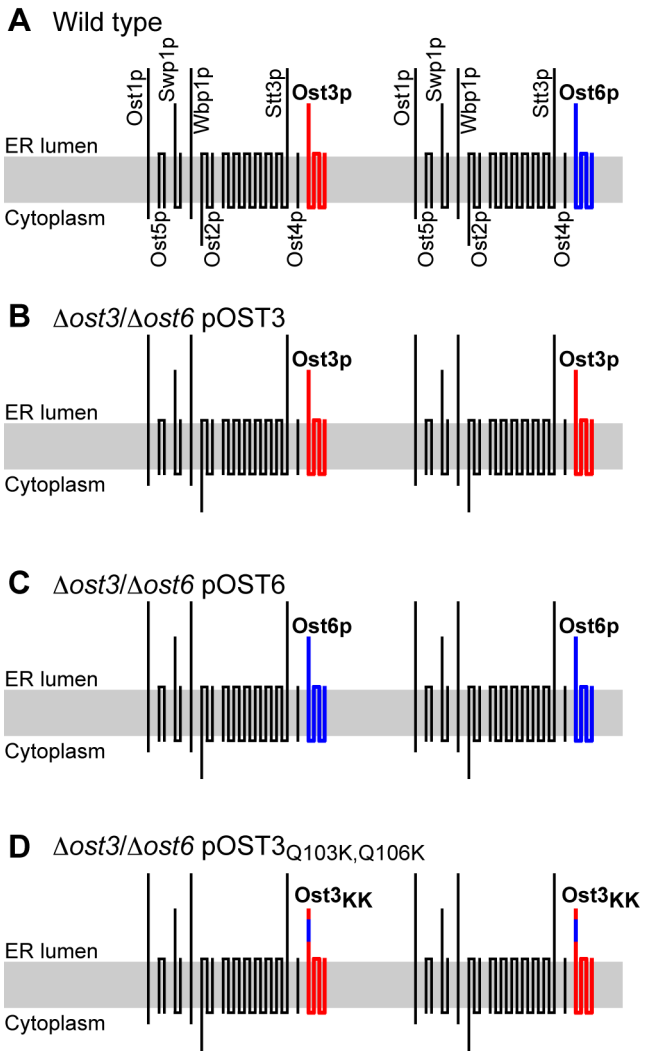

A Wild type
Ost1p
Swp1p
Wbp1p
Stt3p
Ost3p
Ost6p
ER lumen
Cytoplasm
Ost5p
Ost2p
Ost4p
B Δost3/Δost6 pOST3
C Δost3/Δost6 pOST6
D Δost3/Δost6 pOST3Q103K,Q106K
Ost3KK

Figure 2

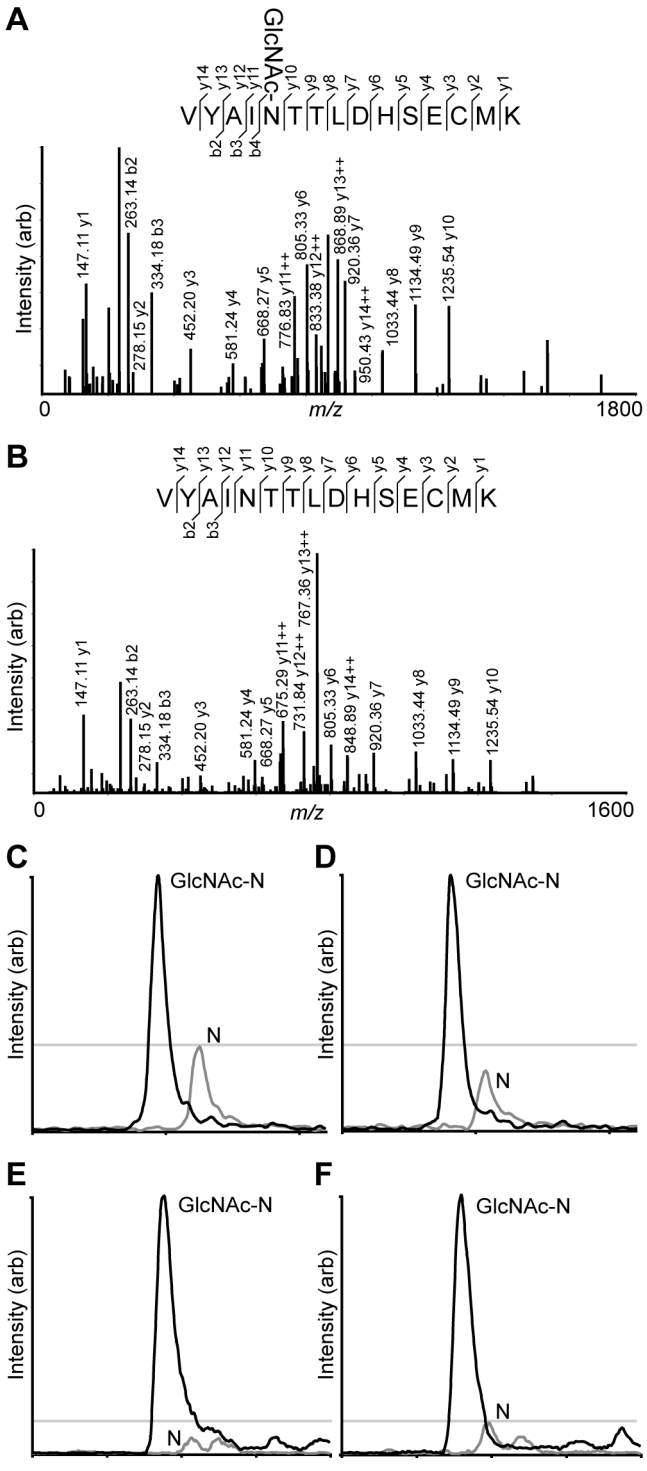

A
GlcNAc-N
V Y A I N T T L D H S E C M K
Intensity (arb)
147.11 y1
263.14 b2
334.18 b3
278.15 y2
452.20 y3
581.24 y4
668.27 y5
776.83 y11++
805.33 y6
833.38 y12++
868.89 y13++
920.36 y7
950.43 y14++
1033.44 y8
1134.49 y9
1235.54 y10
0
m/z
1800
B
V Y A I N T T L D H S E C M K
Intensity (arb)
147.11 y1
263.14 b2
278.15 y2
334.18 b3
452.20 y3
581.24 y4
668.27 y5
675.29 y11++
731.84 y12++
767.36 y13++
805.33 y6
848.89 y14++
920.36 y7
1033.44 y8
1134.49 y9
1235.54 y10
0
m/z
1600
C
GlcNAc-N
N
Intensity (arb)
D
GlcNAc-N
N
Intensity (arb)
E
GlcNAc-N
N
Intensity (arb)
F
GlcNAc-N
N
Intensity (arb)

Figure 3

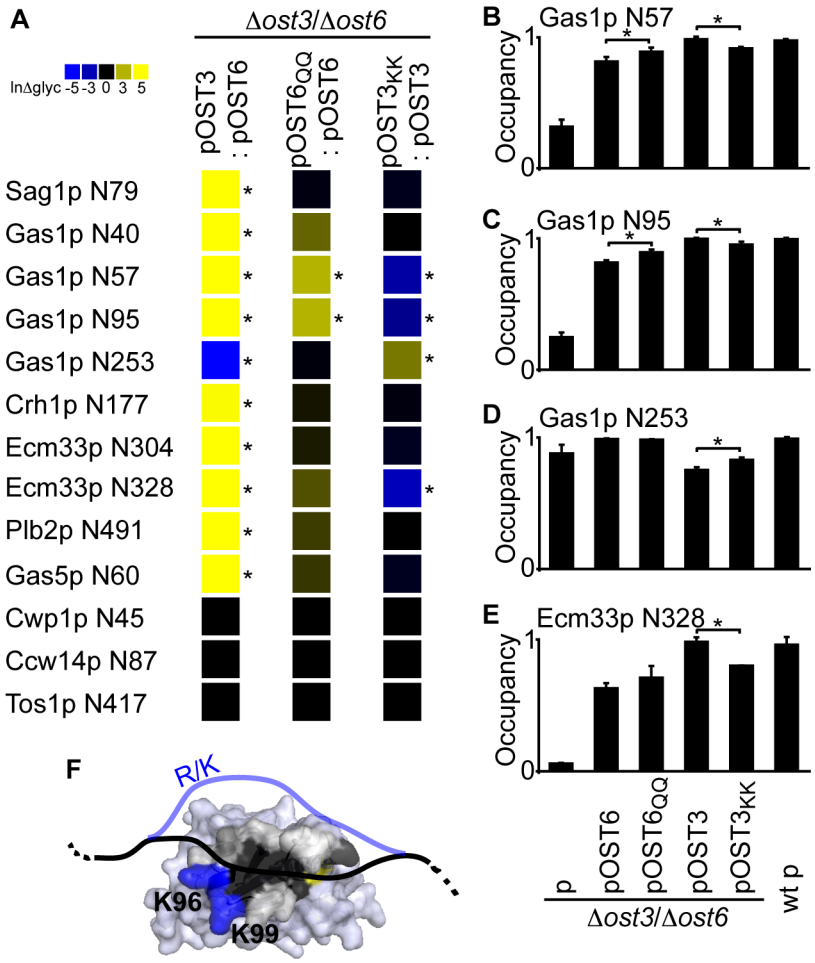

A
Δost3/Δost6
lnΔglyc -5 -3 0 3 5
pOST3 : pOST6
pOST6QQ : pOST6
pOST3KK : pOST3
Sag1p N79
Gas1p N40
Gas1p N57
Gas1p N95
Gas1p N253
Crh1p N177
Ecm33p N304
Ecm33p N328
Plb2p N491
Gas5p N60
Cwp1p N45
Ccw14p N87
Tos1p N417
B
Gas1p N57
C
Gas1p N95
D
Gas1p N253
E
Ecm33p N328
Occupancy
1
0
p
pOST6
pOST6QQ
pOST3
pOST3KK
wt p
Δost3/Δost6
F
R/K
K96
K99

Figure 4

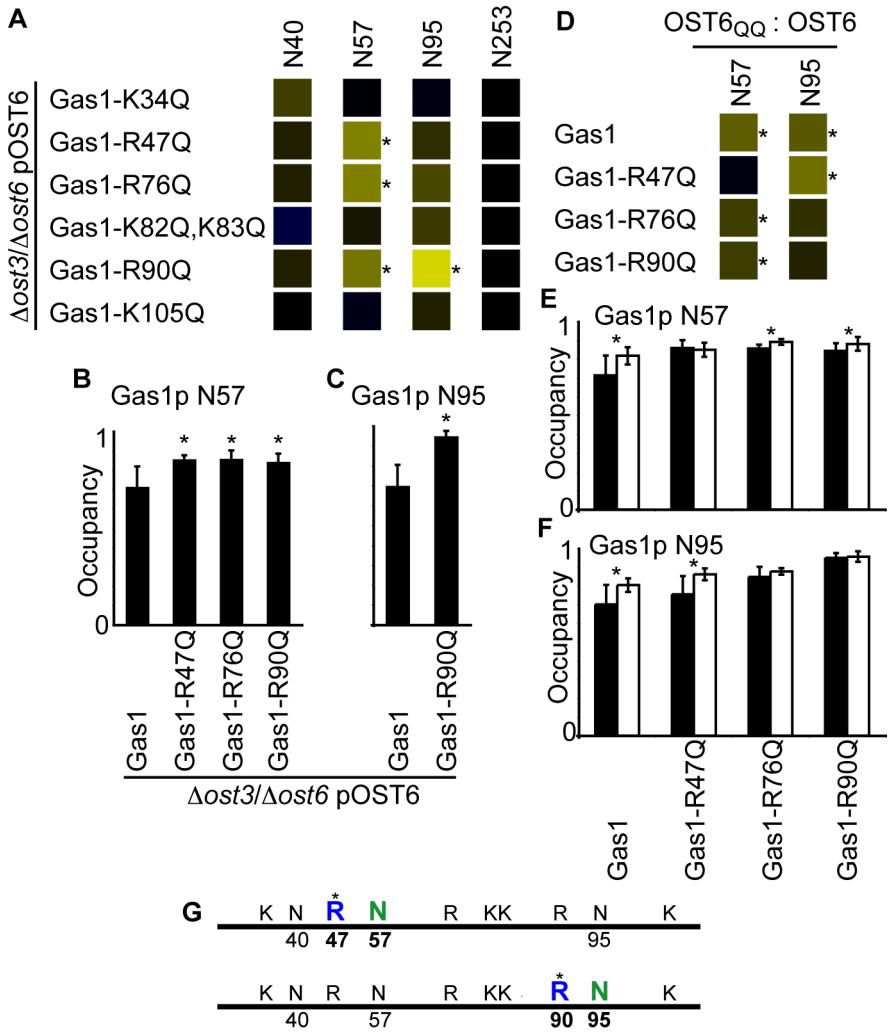

A
N40
N57
N95
N253
Δost3/Δost6 pOST6
Gas1-K34Q
Gas1-R47Q
Gas1-R76Q
Gas1-K82Q,K83Q
Gas1-R90Q
Gas1-K105Q
B
Gas1p N57
Occupancy
1
0
Gas1
Gas1-R47Q
Gas1-R76Q
Gas1-R90Q
C
Gas1p N95
Gas1
Gas1-R90Q
Δost3/Δost6 pOST6
D
OST6QQ : OST6
N57
N95
Gas1
Gas1-R47Q
Gas1-R76Q
Gas1-R90Q
E
Gas1p N57
Occupancy
F
Gas1p N95
Occupancy
Gas1
Gas1-R47Q
Gas1-R76Q
Gas1-R90Q
G
K N R N R KK R N K
40 47 57 95
K N R N R KK R N K
40 57 90 95

Figure 5

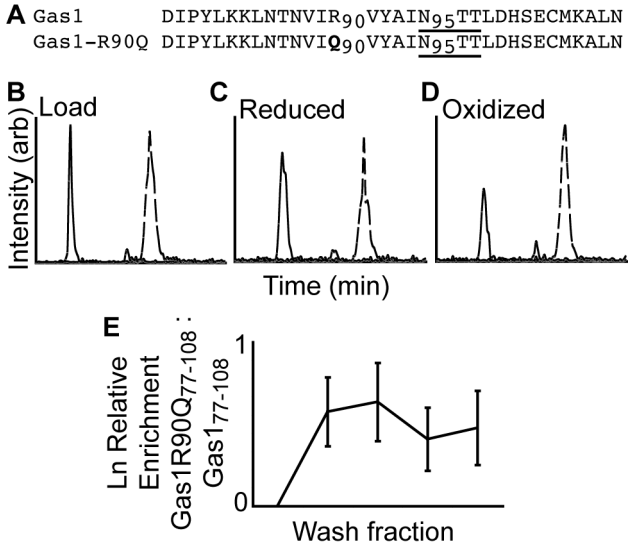

A Gas1 DIPYLKKLNTNVIR90VYAIN95TTLDHSECMKALN
Gas1-R90Q DIPYLKKLNTNVIQ90VYAIN95TTLDHSECMKALN
B Load
C Reduced
D Oxidized
Intensity (arb)
Time (min)
E
Ln Relative Enrichment Gas1R90Q77-108 : Gas177-108
1
0
Wash fraction

Figure 6

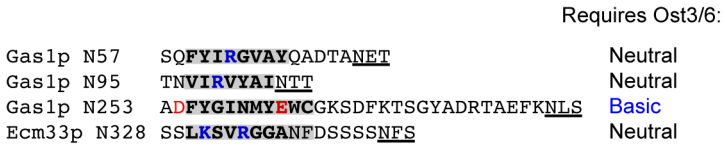
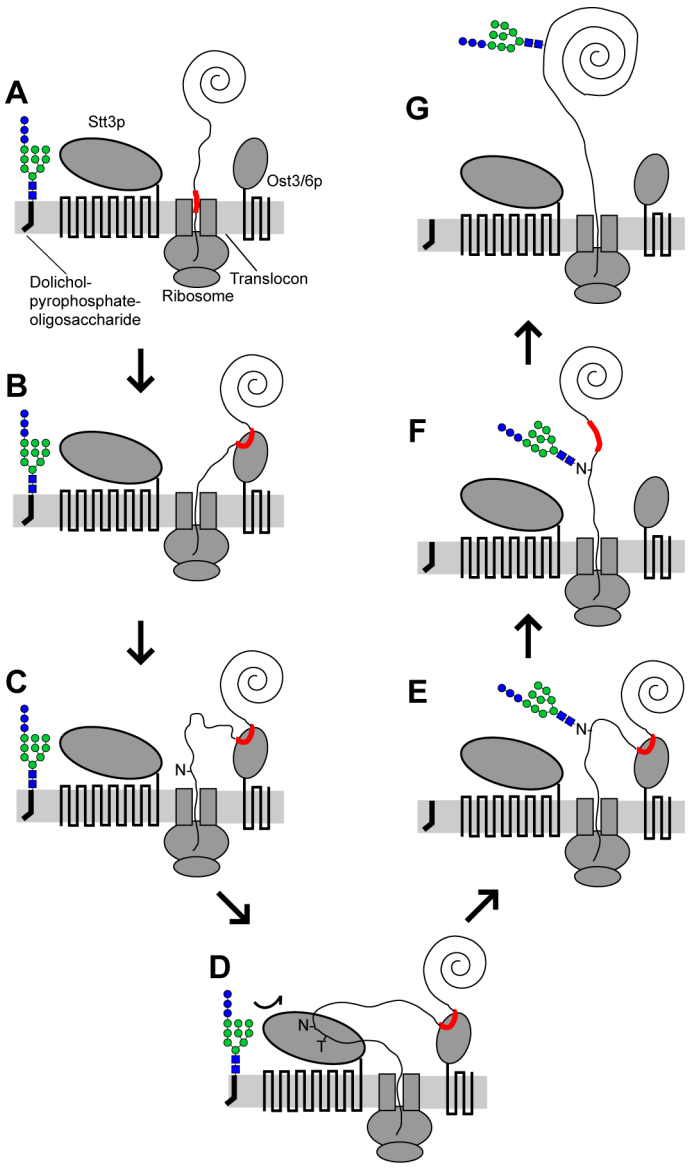

Requires Ost3/6:
Gas1p N57 SQFYIRGVAYQADTANET Neutral
Gas1p N95 TNVIRVYAINTT Neutral
Gas1p N253 ADFYGINMYEWCGKSDFKTSGYADRTAEFKNLS Basic
Ecm33p N328 SSLKSVRGGANFDSSSSNFS Neutral

Figure 7


A
Stt3p
Ost3/6p
Dolichol-
pyrophosphate-
oligosaccharide
Ribosome
Translocon
B
C
N
D
N
T
E
N
F
N
G

2.4 Supplementary Information

Jamaluddin MF, Bailey UM, Schulz BL. Oligosaccharyltransferase subunits bind polypeptide substrate to locally enhance *N*-glycosylation. *Mol Cell Proteomics*. **2014**, 13(12):3286-3293

Table S1. Site-specific N-glycosylation occupancy in wild type and Ost3p/Ost6p variant expressing yeast strains.

Glycosylation site	Strain					
	Wild type	Δ*ost3*/Δ*ost6*				
		pOST3	pOST3 $_{Q103K,Q106K}$	pOST6	pOST6 $_{K96Q,K99Q}$	p
Sag1p N79	0.35±0.03	0.40±0.04*	0.36±0.08	0.04±0.02	0.02±0.00	0.02±0.04
Gas1p N40	1.00±0.00	1.00±0.00*	1.00±0.00	0.94±0.01	0.95±0.01	0.62±0.11
Gas1p N57	0.97±0.01	0.98±0.02*	0.91±0.01*	0.81±0.04	0.89±0.03*	0.32±0.05
Gas1p N95	0.99±0.01	1.00±0.01*	0.95±0.02*	0.81±0.02	0.89±0.02*	0.25±0.03
Gas1p N253	0.99±0.01	0.75±0.02*	0.83±0.02*	0.99±0.00	0.98±0.00	0.88±0.06
Crh1p N177	1.00±0.00	0.99±0.01*	0.99±0.02	0.65±0.02	0.69±0.06	0.27±0.03
Ecm33p N304	0.97±0.04	0.99±0.02*	0.91±0.09	0.36±0.06	0.41±0.02	0.07±0.08
Ecn33p N328	0.96±0.06	0.98±0.04*	0.80±0.00*	0.63±0.04	0.71±0.09	0.06±0.01
Plb2p N491	0.99±0.01	1.00±0.01	1.00±0.01	0.92±0.03	0.95±0.03	0.40±0.18
Gas5p N60	1.00±0.00	1.00±0.00*	0.96±0.07	0.67±0.08	0.77±0.06	0.07±0.02
Cwp1p N45	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
Ccw14p N87	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
Tos1p N417	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00

*, P<0.05 ANOVA comparing occupancy with pOST3 and pOST6, pOST3$_{Q101K,Q103K}$ to pOST3, or pOST6$_{K96Q,K99Q}$ to pOST6. Values show mean. Error shows s.e.m.. n=3.

Table S2. Site-specific N-glycosylation occupancy in Gas1p variants expressed in Δ*ost3*/Δ*ost6* pOST6 yeast.

Glycosylation site	Strain						
	Δ*ost3*/Δ*ost6* pOST6						
	pGAS1	pGAS1 K34Q	pGAS1 R47Q	pGAS1 R76Q	pGAS1 K82Q,K83Q	pGAS1 R90Q	pGAS1 K105Q
Gas1p N40	0.84±0.08	0.91±0.02	n.d.[#]	0.88±0.06	0.80±0.15	0.88±0.09	0.84±0.09
Gas1p N57	0.71±0.10	0.69±0.17	0.85±0.04*	0.85±0.02*	0.73±0.02	0.85±0.05*	0.64±0.10
Gas1p N95	0.70±0.10	0.65±0.16	0.75±0.16	0.84±0.09	0.83±0.03	0.95±0.03*	0.73±0.09
Gas1p N253	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00

*, P<0.05 ANOVA compared to occupancy in wild type Gas1p. Values show mean. Error shows s.e.m.. n=3. [#], n.d. not detected.

Table S3. Site-specific N-glycosylation occupancy in Gas1p variants expressed in Δ*ost3/*Δ*ost6* pOST6 $_{K96Q,K99Q}$ yeast.

Glycosylation site	Strain			
	Δ*Ost3*/Δ*Ost6* Post6$_{K96Q,K99Q}$			
	pGAS1	pGAS1$_{R47Q}$	pGAS1$_{R76Q}$	pGAS1$_{R90Q}$
Gas1p N40	0.96±0.02	0.94±0.02	0.95±0.01	0.96±0.01
Gas1p N57	0.81±0.04*	0.85±0.03	0.89±0.01*	0.88±0.03*
Gas1p N95	0.80±0.03*	0.86±0.03*	0.87±0.01	0.95±0.02
Gas1p N253	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00

*, P<0.05 ANOVA comparing occupancy with pOST6$_{K96Q,K99Q}$ to pOST6. Values show mean. Error shows s.e.m.. n=3.

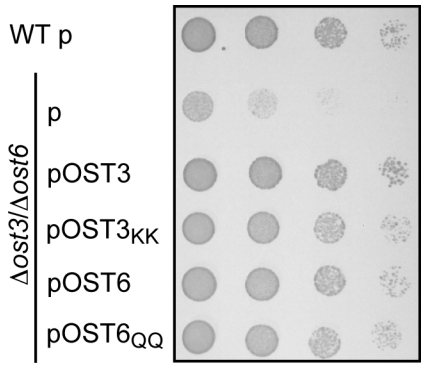


Figure S1. Growth of yeast expressing variant Ost3p or Ost6p. 5-fold serial dilutions of yeast cells from wild type with empty vector (WT p) or *Δost3/Δost6* cells with p, pOST3, $pOST3_{Q101K,Q103K}$, pOST6, or $pOST6_{K96Q,K99Q}$ were grown on minimal media agar for 1 day at 30°C.

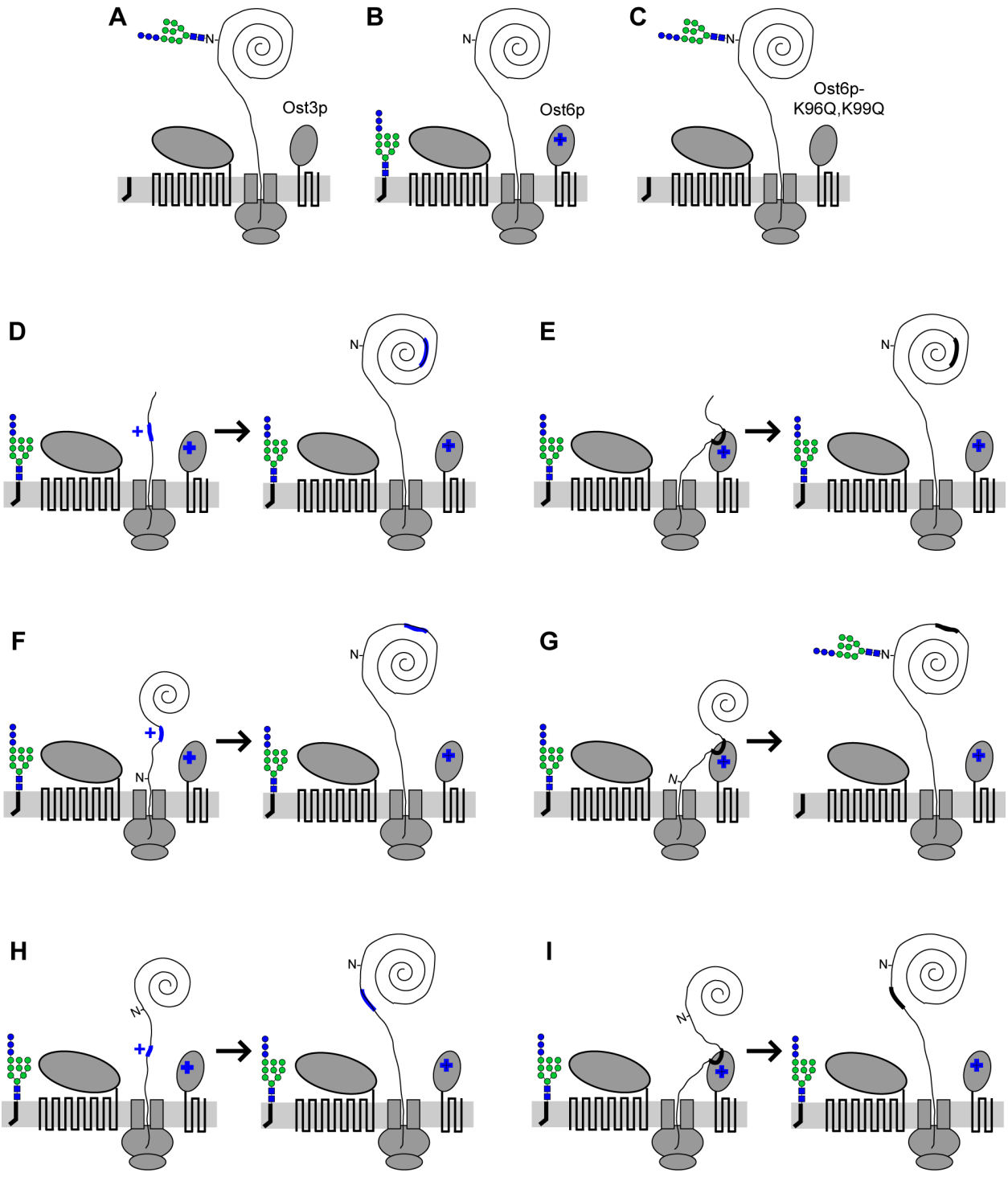


Figure S2. Approach for mapping sites of interaction between protein substrate and Ost6p.

Model substrate protein Gas1p is efficiently glycosylated at e.g. N95 by (**A**) Ost3p or (**C**) Ost6p-$_{K96Q,K00Q}$ but not by (**B**) basic Ost6p. Scanning basic-to-neutral point mutagenesis through Gas1p

maps the site of interaction with Ost6p relevant for glycosylation. (**D**,**F**,**H**) Basic residues (blue) in Gas1p limit glycosylation by preventing interaction with Ost6p. (**E**,**G**,**I**) Basic-to-neutral mutations (black) allow transient binding, but (G) increase glycosylation only at the site of interaction that presents glycosylation-competent N95 to Stt3p (see Fig. 4).

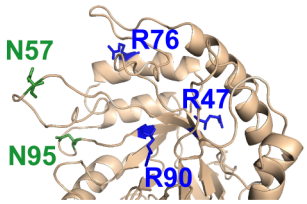


Figure S3. Cartoon of glycosylation sites in mature Gas1p. Gas1p was modelled on the structure of Gas2p (PDB ID 2W61) (20). Green, glycosylation sites N57 and N95. Blue, basic residues R47, R76 and R90.

Chapter 3

Selected reaction monitoring Mass Spectrometry for targeted peptide multi-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p

3.1 Introduction to this publication

In this article, we investigated the *in vitro* interactions between substrate polypeptide and Ost3p and/or Ost6p by peptide affinity chromatography and selected-reaction-monitoring mass spectrometry (SRM-MS). We developed and used SRM-MS for improved relative quantification and sensitivity for detection in conjunction with peptide affinity chromatography. We show that Ost3p and Ost6p bind distinct subsets of peptides from a yeast cell wall glycoprotein substrate, with Ost3p binding peptides rich in aromatic acids and Ost6p binding peptides rich in aliphatic amino acids. Our data supports a model of Ost3p and Ost6p function in which they transiently bind complementary hydrophobic stretches of nascent polypeptide substrate to optimally inhibit protein folding, thereby increasing glycosylation efficiency at nearby asparagine residues.

3.2 Declaration on authorship

In accordance with The University of Queensland, the methods, results and discussion are presented in the form of a publication style in the peer-reviewed journal. This article will be submitted to the journal *Analytical Biochemistry* or *Journal of Chromatography B Analytical Technology Biomedical Life Sciences*. The contributions of all authors are mentioned in Section 3.3

3.3 Article to be submitted

"Selected reaction monitoring Mass Spectrometry for targeted peptide multi-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p"

Jamaluddin MF, Bailey UM, Schulz BL

Draft ready for submission.

Contributions:

This paper describes a detection method using SRM-MS when coupled to peptide affinity chromatography provides significant improvements in sensitivity and quantification that defines the peptide binding characteristics of native yeast Ost3p and Ost6p.

This work presented here was done under the supervision of Dr Benjamin Schulz.

Dr Benjamin Schulz, Dr Maja Bailey and I planned experiments; I performed *in vitro* peptide binding assays; Dr Benjamin Schulz and I performed mass spectrometry; Dr Benjamin Schulz, Dr Maja Bailey and I wrote the manuscript. All the authors contributed to the critical reviewing of the manuscript before the submission.

Article Title: Selected Reaction Monitoring Mass Spectrometry for targeted peptide multi-analyte affinity chromatography defines the peptide-binding activities of yeast Oligosaccharyltransferase subunits Ost3p and Ost6p.

Author(s): M. Fairuz B. Jamaluddin, Ulla-Maja Bailey, and Benjamin L. Schulz*

Affiliations: School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia

***Correspondence to:** Benjamin L. Schulz, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: b.schulz@uq.edu.au

Keywords: *N*-glycosylation; oligosaccharyltransferase; mass spectrometry; protein-peptide interactions

Acknowledgements of financial support: We acknowledge the support of NHMRC project grant 631615 and CDF APP1031542

Abbreviations (Goes to Section 3): ER, endoplasmic reticulum; MALDI-Tof/Tof-MS, Matrix assisted laser desorption/ionization tandem time-of-flight mass spectrometry; SRM-MS, Selected reaction monitoring mass spectrometry; *N*-glycosylation, asparagine-linked glycosylation; OTase, oligosaccharyltransferase

Abstract

Asparagine-linked glycosylation is a common post-translational modification that regulates the structure and function of secretory and membrane proteins in eukaryotes. This process is catalysed by the multiprotein complex oligosaccharyltranferase (OTase). Two isoforms of OTase exist in the yeast *Saccharomyces cerevisiae*, determined by the presence of either of the homologous proteins Ost3p or Ost6p. These accessory OTase protein subunits are not involved in catalysis, but instead regulate the protein substrate specificity of OTase. Ost3p and Ost6p have an ER lumenal thioredoxin-like domain with peptide-binding grooves that can transiently bind stretches of polypeptide. Here, we describe an analytical technique using *in vitro* peptide affinity chromatography and selected-reaction-monitoring mass spectrometry to detect peptides that bind to the peptide-binding grooves of Ost3p and Ost6p. We show that Ost3p and Ost6p bind distinct subsets of peptides from a yeast cell wall glycoprotein substrate, with Ost3p binding peptides rich in aromatic acids and Ost6p binding peptides rich in aliphatic amino acids. Our data supports a model of Ost3p and Ost6p function in which they transiently bind complementary hydrophobic stretches of nascent polypeptide substrate to optimally inhibit protein folding, thereby increasing glycosylation efficiency at nearby asparagine residues.

Text

Introduction

Asparagine (*N*)-linked glycosylation is the most prevalent post-translational modification of secretory proteins in eukaryotes, and also occurs in archaea and some bacteria [1-3]. In eukaryotic cells, oligosaccharyltransferase (OTase) is an integral membrane protein that catalyzes *N*-glycosylation of nascent polypeptides in the lumen of the endoplasmic reticulum (ER) [4]. OTase is physically associated with the translocon complex [5] and transfers an oligosaccaryl moiety $Glucose_3Mannose_9$*N*-$acetylglucosamine_2$ from a dolichol-pyrophosphate donor onto Asn side-chains in emerging polypeptides [4]. The efficiency of glycosylation of Asn residues dramatically increased if they are present in glycosylation "sequons" (Asn-Xaa-Thr/Ser; Xaa≠Pro), the peptide recognition motif of the catalytic site of OTase [6]. *N*-glycosylation is crucial for efficient glycoprotein folding in the ER by locally recruiting the disulfide isomerase ERp57 through the lectins calnexin and calreticulin [7, 8]. Once proteins are correctly folded, they are free to traffic through the Golgi, where further modification and extension of glycan structures can occur [8]. The precise glycan structures present on mature glycoproteins are often vital for their biological functions such as the immune response, embryonic development and cancer [9-12].

Two isoforms of OTase exist in the yeast Saccharomyces cerevisiae, determined by the presence of either of the homologous proteins Ost3p or Ost6p [13-15]. These proteins have an ER lumenal thioredoxin-like domain followed by four transmembrane helices. Efficient glycosylation of selected Asn residues requires Ost3p, while others require Ost6p [13, 14, 16, 17]. Ost3p and Ost6p have thioredoxin-like ER soluble domains with peptide-binding grooves adjacent to their CxxC redox active sites [18]. This groove is only present when the CxxC motif is oxidized, and becomes flexibly disordered upon reduction [18]. The grooves of Ost3p and Ost6p can transiently bind selected peptides with complementary characteristics to these peptide-binding grooves [19]. The groove of Ost6p has a hydrophobic base lined by basic residues, and binds peptides enriched in

hydrophobic and acidic amino acid resides [19]. Furthermore, the peptide-binding activities of Ost3p and Ost6p can be engineered by targeted mutations to their peptide-binding grooves. Point mutations removing the basic residues lining the Ost6p groove abolish peptide binding, while mutations to insert basic residues in place of the natively neutral residues lining the Ost3p groove lead to binding of a hydrophobic and acidic peptide that also binds to wild type Ost6p [19]. These studies used matrix-assisted-laser-desorption-ionization time-of-flight mass spectrometry (MALDI-Tof/Tof-MS) to detect differentially bound peptides [19]. Despite robustly identifying 35 tryptic peptides from a model yeast cell wall glycoprotein, Gas1p, and detecting three peptides that bound to Ost6p and variant Ost3p, this study was unable to identify any peptides that showed significant retention to native Ost3p [19].

Here, we developed and used selected-reaction-monitoring mass spectrometry (SRM-MS) for improved relative quantification and sensitivity for detection in conjunction with peptide affinity chromatography. We used this method to identify peptides from a model substrate protein, Gas1p, that bound to the peptide-binding grooves of both native Ost3p and Ost6p.

Materials and methods

Peptide affinity chromatography

Preparation of reagents for *in vitro* peptide affinity chromatography experiments was performed as previously described [19]. MBP-Ost3p, MBP-Ost6p and MBP-Gas1p were expressed in *E. coli* and purified by amylose-agarose affinity chromatography, and MBP-Gas1p was reduced/alkylated and digested with trypsin. After digestion, residual active trypsin was inactivated by incubation with PMSF. Peptide affinity chromatography was performed as described [19]. Briefly, tryptic peptides from MBP-Gas1p in column buffer (20 mM TrisHCl pH 7.5, 200 mM NaCl, 1 mM EDTA) were applied to amylose-agarose beads containing bound MBP-Ost3p or MBP-Ost6p, with the Ost3p and Ost6p CxxC motifs either oxidized, or reduced by prior incubation with DTT. Peptides were washed from the affinity matrix by sequential gravity flow washing with column buffer, and wash fractions collected. Peptides in the load and wash fractions were then detected by LC-SRM-MS.

Mass spectrometry

Peptides in load and wash fractions were measured by LC-SRM-MS using an ABSCIEX 5500 QTRAP. Peptides were separated by C-18 LC as described [20]. Scheduled SRM transitions were performed according to optimised parameters (Tab 1). The precursor (Q1) to fragment ion (Q3) transitions, used to measure one targeted peptide. (Q1) selects the specific parent mass and (Q3) selects the specific fragment mass. Peptide retention times were spread over the 30 min LC gradient with 120 s SRM detection windows used for acquisition of each transition and a target scan time of 2 s. All LC-SRM-MS experiments were performed in triplicate. The relative enrichment of each peptide in each wash fraction was determined as described [19] by the abundance of the peptide as measured by the integrated SRM peak area as a fraction of the total abundance of all detected peptides in that fraction. Interactions with the peptide-binding groove of Ost3p and Ost6p were determined for each peptide by the log of the relative abundance of each peptide in each wash

fraction between the oxidized and reduced versions of Ost3p or Ost6p using a two-sided paired student's t-test with $P < 0.005$. Peptide characteristics were determined using ProtParam [21].

Results and Discussion

The Ost3p and Ost6p accessory subunits of yeast OTase assist efficient *N*-glycosylation at distinct subset of glycosylated asparagines [16]. A model of Ost3/6p function proposes that these accessory proteins assist *N*-glycosylation by transiently binding stretches of nascent polypeptide complementary to the characteristics of their peptide-binding grooves [18]. In this study we have performed peptide affinity chromatography coupled to LC-SRM-MS to identify the characteristics of peptides that bind to both native Ost3p and Ost6p.

Previous similar in vitro peptide binding experiments identified stretches of polypeptide that interacted with the peptide-binding grooves of Ost6p and variant Ost3p [19]. These studies used a model glycoprotein substrate of Ost3/6p, the yeast cell wall glycoprotein Gas1p. Gas1p was efficiently expressed and purified as MBP-fusion protein in E.coli in order to obtain unglycosylated polypeptide similar to nascent polypeptide entering the ER lumen, which is the physiological substrate of OTase. Purified unglycosylated MBP-Gas1p was digested with Trypsin and applied to resin containing bound MBP-Ost3p or MBP-Ost6p, with the active-site CxxC motif either oxidized, or reduced by pre-incubation with DTT. The peptide-binding groove is only present when the CxxC motif is oxidized, and peptides do not bind when this motif is reduced [18].

We performed peptide affinity chromatography as previously described [19], and then measured the MBP-Gas1p tryptic peptides present in the load and wash fractions by LC-SRM-MS. This rapid, sensitive detection and relative quantification proved superior to the previously utilized detection with MALDI-Tof/Tof-MS [19]. Of the 38 MBP-Gas1p tryptic peptides robustly detected, 14 peptides were identified that were retained by affinity columns with oxidized, but not reduced,

MBP-Ost6p, indicating that they bound at the peptide-binding groove of Ost6p (Tab. 2). Comparison of the biophysical characteristics of peptides that bound or did not bind to the peptide-binding groove of Ost6p were in agreement with the previously reported ability of Ost6p to specifically bind peptides enriched in aliphatic residues (Fig. 1B) [19]. We then tested if LC-SRM-MS improved detection for peptide affinity chromatography of MBP-Gas1p peptides interacting with the Ost3p. Indeed, our LC-SRM-MS approach successfully identified six peptides that were retained by affinity matrix with oxidized, but not reduced, MBP-Ost3p, indicating that they bound at the Ost3p peptide-binding groove. These peptides that bound to Ost3p were significantly enriched in aromatic amino acids (Fig 1D). Notably, the aromatic amino acids in these peptides tended to be clustered, fitting the approximate dimensions of the Ost3/6p peptide binding grooves [18, 19].

Ost3p and Ost6p bound distinct sets of tryptic peptides from MBP-Gas1p, as Ost3p bound peptides enriched with aromatic residues, while Ost6p bound peptides enriched with aliphatic residues (Tab. 2, Fig. 1). The crystal structure of Ost6p shows that it possesses a peptide-binding groove with a base of aliphatic amino acids (Fig. 2A,B) [18]. Sequence alignment of Ost3p and Ost6p shows that Ost3p has an aromatic amino acid Phe102 in place of the aliphatic Val95 in Ost6p (Fig. 2C). Thus, binding of peptides enriched with aliphatic residues by the peptide-binding groove of Ost6 correlates with the aliphatic amino acids forming the base of the groove in Ost6p, while in contrast, Ost3p binds peptides rich in aromatic amino acids, correlating with the presence of Phe in its peptide-binding groove. That is, the peptides that bound to Ost3p and Ost6p had different physical characteristics complementary to the respective characteristics of the peptide-binding grooves of Ost3p and Ost6p. This is consistent with the proposed model of Ost3/6p activity, in that different stretches of nascent polypeptide would be bound by OTase isoforms containing Ost3p or Ost6p, which would enhance glycosylation efficiency of nearby sequons in an OTase isoform-dependent manner [16, 18]. Transient tethering of hydrophobic sections of nascent polypeptide would

efficiently inhibit local protein folding, as such hydrophobic sections are typically internal to mature folded protein domains. This correlates with the requirement of the catalytic site of OTase for unfolded protein acceptor substrate [6]. The distinct peptide binding activities of Ost3p and Ost6p would broaden the range of this protein folding inhibition, with binding of hydrophobic stretches of nascent polypeptide enriched with aliphatic amino acids by Ost6p and aromatic acids by Ost3p. This apparent duplication and diversification of OTase substrate specificity is similar to the different protein acceptor specificities based on local charge of the single protein OTases in *Trypanosoma brucei* [22, 23], and based on ill-defined characteristics in *Leishmania major* [24].

In conclusion, we have made use of sensitive LC-SRM-MS detection for relative quantification following multi-peptide affinity chromatography to define the peptide binding characteristics of native yeast Ost3p and Ost6p. Use of LC-SRM-MS as a detection method when coupled to peptide affinity chromatography provided significant improvements in sensitivity and quantification, and as such similar methods may be generically applicable for other peptide or multi-analyte affinity chromatography.

References

1. Calo, D., L. Kaminski, and J. Eichler, *Protein glycosylation in Archaea: sweet and extreme.* Glycobiology, 2010. **20**(9): p. 1065-76.

2. Lehle, L., S. Strahl, and W. Tanner, *Protein glycosylation, conserved from yeast to man: a model organism helps elucidate congenital human diseases.* Angew Chem Int Ed Engl, 2006. **45**(41): p. 6802-18.

3. Nothaft, H. and C.M. Szymanski, *Protein glycosylation in bacteria: sweeter than ever.* Nat Rev Microbiol, 2010. **8**(11): p. 765-78.

4. Kelleher, D.J. and R. Gilmore, *An evolving view of the eukaryotic oligosaccharyltransferase.* Glycobiology, 2006. **16**(4): p. 47R-62R.

5. Harada, Y., et al., *Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site.* Proceedings of the National Academy of Sciences, 2009. **106**(17): p. 6945-6949.

6. Lizak, C., et al., *X-ray structure of a bacterial oligosaccharyltransferase.* Nature, 2011. **474**(7351): p. 350-5.

7. Dempski, R.E. and B. Imperiali, *Oligosaccharyl transferase: gatekeeper to the secretory pathway.* Current Opinion in Chemical Biology, 2002. **6**(6): p. 844-850.

8. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annual Review of Biochemistry, 2004. **73**: p. 1019-1049.

9. Marth, J.D. and P.K. Grewal, *Mammalian glycosylation in immunity.* Nat Rev Immunol, 2008. **8**(11): p. 874-887.

10. Ohtsubo, K. and J.D. Marth, *Glycosylation in cellular mechanisms of health and disease.* Cell, 2006. **126**(5): p. 855-867.

11. Schulz, B.L., W. Laroy, and N. Callewaert, *Clinical laboratory testing in human medicine based on the detection of glycoconjugates.* Curr Mol Med, 2007. **7**(4): p. 397-416.

12. Zhao, Y.Y., et al., *Functional roles of N-glycans in cell signaling and cell adhesion in cancer.* Cancer Sci, 2008. **99**(7): p. 1304-10.

13. Schwarz, M., R. Knauer, and L. Lehle, *Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit.* FEBS Lett, 2005. **579**(29): p. 6564-6568.

14. Spirig, U., et al., *The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast.* Glycobiology, 2005. **15**(12): p. 1396-1406.

15. Yan, A. and W.J. Lennarz, *Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons.* Glycobiology, 2005. **15**(12): p. 1407-1415.

16. Schulz, B.L. and M. Aebi, *Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency*. Molecular & Cellular Proteomics, 2009. **8**(2): p. 357-364.

17. Yan, A. and W.J. Lennarz, *Unraveling the mechanism of protein N-glycosylation.* J Biol Chem, 2005. **280**(5): p. 3121-4.

18. Schulz, B.L., et al., *Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency*. Proceedings of the National Academy of Sciences, 2009.

19. Jamaluddin, M.F.B., et al., *Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p.* Protein Science, 2011. **20**(5): p. 849-855.

20. Ku, S.C., et al., *The pilin O-glycosylation pathway of pathogenic Neisseria is a general system that glycosylates AniA, an outer membrane nitrite reductase.* Biochem Biophys Res Commun, 2009. **378**(1): p. 84-89.

21. Gasteiger, E., et al., *Protein Identification and Analysis Tools on the ExPASy Server*. 2005. p. 571-607.

22. Izquierdo, L., et al., *Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. EMBO J*, 2009. **28**(17): p. 2650-2661.

23. Izquierdo, L., A. Mehlert, and M.A. Ferguson, *The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases.* Glycobiology, 2012. **22**(5): p. 696-703.

24. Nasab, F.P., et al., *All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae.* Mol Biol Cell, 2008. **19**(9): p. 3758-68.

Table 1: SRM transitions and instrument parameters.

Q1 mass (Da)	Q3 mass (Da)	Time (min)	Peptide sequence	CE (volts)
529.31	732.37	14.2	MBP-Gas1.LVIWINGDK.2/y6	23.000
504.26	787.43	10	MBP-Gas1.GYNGLAEVGK.2/y8	22.000
462.74	625.29	12.4	MBP-Gas1.VTVEHPDK.2/y5	25.000
738.36	1057.56	16.6	MBP-Gas1.FPQVAATGDGPDIIFWAHDR.3/y8	41.000
883.95	1143.63	15.5	MBP-Gas1.FGGYAQSGLLAEITPDK.2/y11	44.000
634.33	991.5	16.3	MBP-Gas1.LYPFTWDAVR.2/y8	28.000
946.05	1149.65	22.1	MBP-Gas1.LIAYPIAVEALSLIYNK.2/y10	47.000
447.26	552.31	11.7	MBP-Gas1.DLLPNPPK.2/y5	25.000
601.31	785.44	13.7	MBP-Gas1.TWEEIPALDK.2/y7	31.000
984.15	1222.65	22.1	MBP-Gas1.SALMFNLQEPYFTWPLIAADGGYAFK.3/y12	48.000
595.36	948.58	20	MBP-Gas1.AGLTFLVDLIK.2/y8	26.000
949.42	1113.56	13.2	MBP-Gas1.HMNADTDYSIAEAAFNK.2/y10	47.000
727.01	1021.49	16.4	MBP-Gas1.GETAMTINGPWAWSNIDTSK.3/y9	35.000
669.38	805.48	14.9	MBP-Gas1.VNYGVTVLPTFK.2/y7	29.000
713.72	1029.53	14.4	MBP-Gas1.GQPSKPFVGVLSAGINAASPNK.3/y11	35.000
699.68	730.41	17	MBP-Gas1.EFLENYLLTDEGLEAVNK.3/y7	34.000
506.31	768.5	10.4	MBP-Gas1.DKPLGAVALK.2/y8	27.000
668.82	957.5	9.1	MBP-Gas1.SYEEELAKDPR.2/y8	39.000
704.01	999.5	19.5	MBP-Gas1.GEIMPNIPQMSAFWYAVR.3/y8	34.000
480.27	688.37	8.5	MBP-Gas1.TAVINAASGR.2/y7	26.000
625.98	926.53	14.3	MBP-Gas1.ISEFGSDDVPAIEVVGNK.3/y9	30.000
821.89	984.49	14.5	MBP-Gas1.FFYSNNGSQFYIR.2/y8	41.000
773.59	1210.55	12.8	MBP-Gas1.GVAYQADTANETSGSTVNDPLANYESC[PPa]SR.4/y10	37.000
374.72	520.31	12.3	MBP-Gas1.DIPYLK.2/y4	21.000
415.25	602.36	9.4	MBP-Gas1.LNTNVIR.2/y5	28.000
599.28	805.33	12.1	MBP-Gas1.VYAINTTLDHSEC[PPa]MK.3/y6	39.000
763.07	1013.57	18	MBP-Gas1.ALNDADIYVIADLAAPATSINR.3/y10	47.000
850.89	985.5	16.8	MBP-Gas1.DDPTWTVDLFNSYK.2/y8	42.000
784.35	951.37	12.9	MBP-Gas1.IPVGYSSNDDEDTR.2/y8	35.000
784.35	1038.4	8.9	MBP-Gas1.IPVGYSSNDDEDTR.2/y9	45.000
775.81	893.37	12.5	MBP-Gas1.MTDYFAC[PPa]GDDDVK.2/y8	39.000
884.38	1101.45	16.2	MBP-Gas1.ADFYGINMYEWC[PPa]GK.2/y8	39.000
748.36	1109.5	17.7	MBP-Gas1.NLSIPVFFSEYGC[PPa]NEVTPR.3/y9	36.000
640.33	760.38	12.7	MBP-Gas1.YGLVSIDGNDVK.2/y7	33.000
830.38	954.45	12.5	MBP-Gas1.TLDDFNNYSSEINK.2/y8	42.000
459.75	718.37	15.6	MBP-Gas1.ISPTSANTK.2/y7	25.000
604.28	753.36	11.8	MBP-Gas1.YGAYSFC[PPa]TPK.2/y6	32.000
832.41	1059.52	16.6	MBP-Gas1.EQLSFVMNLYYEK.2/y8	37.000
865.4	1167.57	12.6	MBP-Gas1.SDC[PPa]SFSGSATLQTATTQASC[PPa]SSALK.3/y11	42.000
700.62	1001.44	16.9	MBP-Gas1.SNSGSSGSSSSSSSSSASSSSSSK.3/y11	34.000

Table 2. Binding of tryptic peptides from MBP-Gas1p to Ost3p- or Ost6p-affinity chromatography with LC-SRM-MS detection.

Peptide sequence	Log retention (Oxidized/reduced)	
	Ost3p	Ost6p
LVIWINGDK	-0.77 ± 0.01	**0.92 ± 0.01***
GYNGLAEVGK	-0.09 ± 0.43	**2.61 ± 0.02**
VTVEHPDK	-0.38 ± 0.42	-0.25 ± 0.39
FPQVAATGDGPDIIFWAHDR	-0.13 ± 0.33	-0.22 ± 0.02
FGGYAQSGLLAEITPDK	-0.62 ± 0.01	-0.45 ± 0.00
LYPFTWDAVR	-0.82 ± 0.54	**0.34 ± 0.01**
LIAYPIAVEALSLIYNK	**0.52 ± 0.00**	-1.56 ± 0.00
DLLPNPPK	0.28 ± 0.04	**2.51 ± 0.02**
TWEEIPALDK	-0.34 ± 0.86	**1.76 ± 0.00**
SALMFNLQEPYFTWPLIAADGGYAFK	**1.09 ± 0.01**	-3.36 ± 0.00
AGLTFLVDLIK	0.23 ± 0.69	-0.72 ± 0.00
HMNADTDYSIAEAAFNK	-0.05 ± 0.44	0.46 ± 0.16
GETAMTINGPWAWSNIDTSK	**1.10 ± 0.02**	-0.59 ± 0.10
VNYGVTVLPTFK	-0.95 ± 0.05	**0.34 ± 0.04**
GQPSKPFVGVLSAGINAASPNK	-0.22 ± 0.34	-0.66 ± 0.00
EFLENYLLTDEGLEAVNK	0.30 ± 0.89	0.02 ± 0.61
DKPLGAVALK	-0.15 ± 0.62	**2.56 ± 0.01**
SYEEELAKDPR	-0.61 ± 0.55	0.66 ± 0.17
GEIMPNIPQMSAFWYAVR	**0.80 ± 0.01**	-2.63 ± 0.21
TAVINAASGR	-0.50 ± 0.59	**1.88 ± 0.02**
ISEFGSDDVPAIEVVGNK	-0.60 ± 0.43	**1.71 ± 0.00**
FFYSNNGSQFYIR	-0.04 ± 0.78	-0.68 ± 0.01
GVAYQADTANETSGSTVNDPLANYESC[PPa]SR	0.78 ± 0.18	-0.71 ± 0.09
DIPYLK	0.23 ± 0.04	**1.80 ± 0.01**
LNTNVIR	-0.12 ± 0.98	**1.60 ± 0.03**
VYAINTTLDHSEC[PPa]MK	-0.80 ± 0.34	0.25 ± 0.25
ALNDADIYVIADLAAPATSINR	-0.37 ± 0.93	-0.35 ± 0.00
DDPTWTVDLFNSYK	-0.29 ± 0.25	**0.22 ± 0.01**
MTDYFAC[PPa]GDDDVK	0.19 ± 0.10	-0.03 ± 0.69
ADFYGINMYEWC[PPa]GK	**2.77 ± 0.00**	-0.18 ± 0.64
NLSIPVFFSEYGC[PPa]NEVTPR	0.59 ± 0.76	-0.36 ± 0.00
YGLVSIDGNDVK	-0.30 ± 0.22	**2.67 ± 0.01**
TLDDFNNYSSEINK	0.72 ± 0.18	**2.60 ± 0.00**
ISPTSANTK	-0.44 ± 0.29	1.38 ± 0.43
YGAYSFC[PPa]TPK	-0.33 ± 0.92	0.01 ± 0.54
EQLSFVMNLYYEK	**1.62 ± 0.00**	-1.19 ± 0.19
SDC[PPa]SFSGSATLQTATTQASC[PPa]SSALK	0.52 ± 0.87	-0.08 ± 0.85
SNSGSSGSSSSSSSSSASSSSSSK	-0.70 ± 0.00	0.12 ± 0.54

* Bold, statistically significant binding to oxidized versus reduced, T-test, $p<0.005$

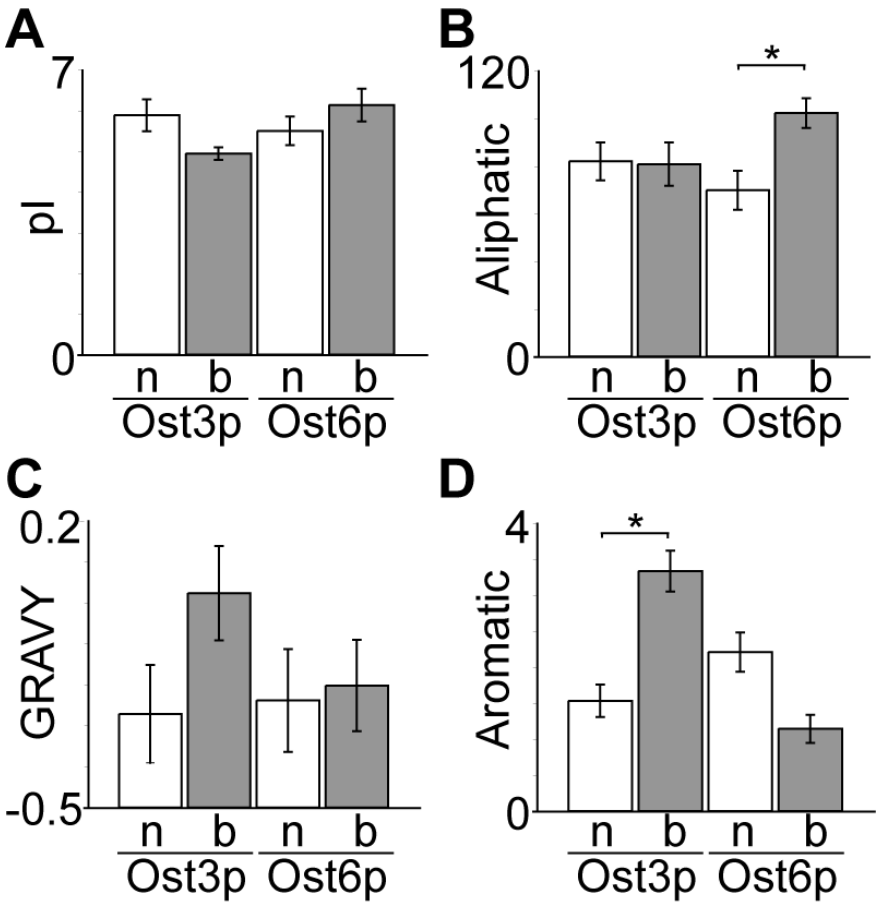


Figure 1. Characteristics of peptides bound (b) or not bound (n) by the peptide-binding groove of Ost3p or Ost6p. (A) pI, (B) aliphatic index, (C) Grand average of hydropathy (GRAVY), (D) aromatic residues. Values show mean. Error bars show standard error. *, Statistically significant, T-test $P<0.005$.

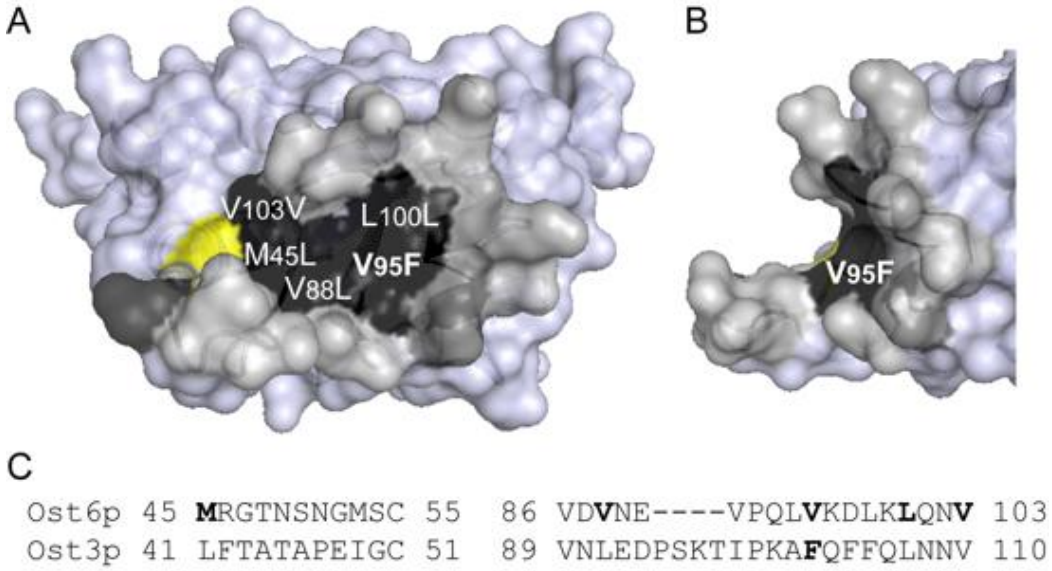


Figure 2. Surface representation of the peptide-binding groove of the ER lumenal domain of Ost6p. (A) top and (B) side view. Hydrophobic amino acids forming the base of the peptide-binding groove are black. Amino acid residues are labelled with their position in Ost6p and the equivalent amino acid at the aligned position in Ost3p. (C) Alignment of amino acid residues forming the peptide-binding groove in Ost3p and Ost6p. Residues labelled in (a) and (b) are bold.

Chapter 4

Analysis of Congenital Disorder of Glycosylation-Id in a Yeast Model System Shows Diverse Site-Specific Under-glycosylation of Glycoproteins

4.1 Introduction to this publication

This chapter was published in *Journal of Proteome Research* as an original article. In this study, we developed an LC-ESI-MS/MS based analytical method to detect peptides/glycopeptides and measure site-specific glycosylation occupancy in *alg3* mutant and wild type yeast *Sacchromyces cervisiae* strains. We performed deglycosylation with peptide-*N*-glycosidase F (PNGase F) followed by AspN digestion and observed an increased number of glycosylation sites in both wild type and Δ*alg3* cells, when compared with PNGase F and trypsin digest. Of the 26 new sites we found in this study, 22 were identified using the combination of PNGase F and AspN. This approach will be useful in site-specific glycosylation analysis in many model systems and clinical application.

4.2 Declaration on authorship

In accordance with The University of Queensland, the methods, results and discussion are presented in the form of a published article in the peer-reviewed journal. This article was published in Journal of Proteome Research. The contributions of all authors are mentioned in Section 4.3

4.3 Published peer-reviewed article

"Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins"

Bailey UM, **Jamaluddin MF**, Schulz BL. Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J. Proteome Res*. **2012**, 11(11):5376-5383.

Contributions:

This paper describes an improved analytical method for site-specific glycosylation analysis using PNGase F to label glycosylation sites with an asparagines-aspartate conversion that creates a new endoproteinase AspN cleavage site, followed by proteolytic digestion, and detection of peptides and glycopeptides by LC-ESI-MS/MS.

This work presented here was done under the supervision of Dr Benjamin Schulz.

Dr Benjamin Schulz, Dr Maja Bailey and I planned experiments; Dr Maja Bailey and I performed *in vivo* yeast experiments; Dr Benjamin Schulz and Dr Maja Bailey performed mass spectrometry; Dr Benjamin Schulz, Dr Maja Bailey and I wrote the manuscript. All the authors contributed to the critical reviewing of the manuscript before the submission.

Journal of proteome research

Article

pubs.acs.org/jpr

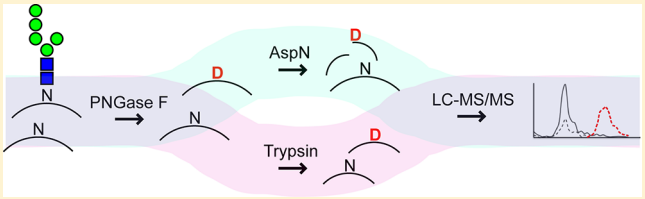
Analysis of Congenital Disorder of Glycosylation-Id in a Yeast Model System Shows Diverse Site-Specific Under-glycosylation of Glycoproteins

Ulla-Maja Bailey, Muhammad Fairuz Jamaluddin, and Benjamin L. Schulz*

School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia

Supporting Information

ABSTRACT: Asparagine-linked glycosylation is a common post-translational modification of proteins in eukaryotes. Mutations in the human ALG3 gene cause changed levels and altered glycan structures on mature glycoproteins and are the cause of a severe congenital disorder of glycosylation (CDG-Id). Diverse glycoproteins are also under-glycosylated in *Saccharomyces cerevisae alg3* mutants. Here we analyzed site-specific glycosylation occupancy in this yeast model system using peptide-N-glycosidase F to label glycosylation sites with an asparagine-aspartate conversion that creates a new endoproteinase AspN cleavage site, followed by proteolytic digestion, and detection of peptides and glycopeptides by LC−ESI−MS/MS. We used this analytical method to identify and measure site-specific glycosylation occupancy in *alg3* mutant and wild type yeast strains. We found decreased site-specific N-glycosylation occupancy in the *alg3* knockout strain preferentially at Asn-Xaa-Ser sequences located in secondary structural elements, features previously associated with poor glycosylation efficiency. Furthermore, we identified 26 previously experimentally unverified glycosylation sites. Our results provide insights into the underlying mechanisms of disease in CDG-Id, and our methodology will be useful in site-specific glycosylation analysis in many model systems and clinical applications.



KEYWORDS: *ALG, alg3, N-glycosylation, oligosaccharyltransferase, mass spectrometry, glycosylation occupancy, PNGase F, CDG*

INTRODUCTION

Protein N-glycosylation is a common and essential co- and post-translocational modification of membrane and secreted proteins that in eukaryotes is catalyzed by the multiprotein complex enzyme oligosaccharyltransferase (OTase).[1] The oligosaccharide $Glc_3Man_9GlcNAc_2$ is assembled on the lipid carrier dolicholpyrophosphate and transferred *en bloc* by OTase onto selected asparagines, primarily in N-glycosylation sequons (NX(S/T); X ≠ P) as nascent polypeptide is translocated into the endoplasmic reticulum (ER).[1,2] Further extension and modification of the glycan structures on mature proteins occurs in the Golgi.[3−5] N-Glycans assist protein folding in the ER, and the specific glycan structures on mature glycoproteins modulate protein function in developmental programs, during immune responses, in cell−cell interactions, and in many other cellular processes.[6−8]

A number of the genes that encode for enzymes in the N-glycosylation pathway have been shown to be involved in human health and disease.[6,9] Many studies have revealed that alterations in glycan expression are associated as causative or incidental factors in relatively rare congenital disorders of glycosylation (CDG), widespread acquired diseases, and cancer.[6,7,9−11] CDGs are conceptually divided into CDG-I, where the genetic defect is in the ER in the N-glycosylation pathway before addition of glycan to protein, and CDG-II, where the defect is in the Golgi. Thus, CDG-I results in inefficient glycosylation of asparagines, while CDG-II results in changes to the glycan structures on mature glycoproteins.[6]

Under-glycosylation is the underlying cause of disease in CDG-I, resulting in an increase in misfolded glycoproteins and alterations in their functions through reduced glycosylation of particular asparagines. However, with over half of all proteins predicted to be glycosylated, the sequence and structural diversity of protein acceptor sites is extraordinarily large.[12] It is currently not clear which particular subset of asparagines are under-glycosylated in CDG-I. However, a study using targeted SRM-MS showed that transferrin isolated from the blood of CDG-I patients had lower glycosylation occupancy specifically at its second N-glycosylation site.[13] Additionally, several familial mutations in the homologous OTase subunits MagT1 and Tusc3 have been linked to a form of CDG-I leading to non-syndromic mental retardation.[14−16] Studies using a manipulable yeast genetic system showed that the yeast homologues of these proteins are required for efficient site-specific N-glycosylation of selected asparagines.[17,18] Other CDGs have also been extensively studied in yeast. *Saccharomyces cerevisiae alg* mutants (*alg*, asparagine-linked glycosylation) are defective in the biosynthesis of the dolichol-linked $Glc_3Man_9GlcNAc_2$,[19,20] and the mutations in the human *alg* homologues causing

Received: July 3, 2012
Published: October 5, 2012

different types of CDG can be functionally studied by heterologous expression in yeast.[6] The ALG3 gene is highly conserved in eukaryotes, and its mutation also results in a similar phenotype in *Arabidobsis thalina*.[21,22] Knockout mutants of these *ALG* genes in *S. cerevisiae* are generally viable and have also been used to characterize the pathway of N-glycosylation in the ER.[19]

Glycosylation occupancy can be measured semiquantitatively on many proteins using MS-based methods.[17,18,23−25] Traditionally in these approaches in yeast, Endo H is used as a deglycosylation agent followed by protease treatment and MS analysis to identify glycosylated and non-glycosylated peptides. Enrichment of glycoproteins rather than glycopeptides allows detection of both glycosylated and non-glycosylated versions of the same peptide. This allows direct relative quantification of site-specific glycosylation occupancy, although it entails some reduction in the enrichment efficiency obtainable through glycopeptide enrichment.[25] Endo H cleaves all N-glycans of wild type yeast, which are all high-mannose structures, and conveniently leaves previously glycosylated asparagines tagged with a single *N*-acetylglucosamine (GlcNAc).[17] Glycan release with Endo H therefore provides a clear distinction between the glycosylated and unglycosylated versions of the same peptide with a Δmass of 203.08 Da. However, in *S. cerevisiae* Δ*alg3* mutants the incomplete precursor glycan $Man_5GlcNAc_2$-PP-Dol accumulates and Endo H resistant oligosaccharides are transferred to proteins.[26]

In this study we chose a yeast Δ*alg3* strain as a model system to develop an MS-based method that can measure the occupancy of specific glycosylation sites in different proteins in situations where Endo H cannot be used.

EXPERIMENTAL SECTION

Yeast Strains

Yeast strains used were BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and the Δ*alg3* mutant derivative thereof (Open Biosystems). Cells were grown to mid-log phase at 30 °C in YPD (2% Bacto peptone, 1% yeast extract, and 2% glucose).

Cell Wall Protein Sample Preparation

Triplicates of 50 mL of cells grown to mid-log phase were harvested and lysed with agitation for 1 h at 4 °C using glass beads (Sigma) in 50 mM Tris HCl, pH 7.5, 1 mM PMSF and 1x Complete protease inhibitor cocktail (Roche). Based on previously reported methods,[17,27] covalently linked cell wall material was pelleted by centrifugation at 18000 rcf for 1 min and washed 4 times with ice-cold 50 mM Tris-HCl pH 7.5. The pellet was resuspended in 50 mM Tris-HCl pH 8.0, 2% SDS, 7 M urea, and 2 M thiourea. Cysteines were reduced and alkylated by addition of dithiothretiol to 10 mM and incubation at 30 °C with agitation for 30 min, followed by addition of acrylamide to 50 mM and further incubation with agitation at 30 °C for 1 h. Noncovalently linked proteins were removed by washing the 18000 rcf pellet five times with 1 mL of 50 mM Tris-HCl, pH 8, 2% SDS, 7 M urea, and 2 M thiourea, followed by five washes with 2% SDS. The pellet was resuspended in 100 μL of 1% NP40 and 1x G7 buffer (New England Biolabs) and aliquoted into 50 μL followed by addition of 500 units of PNGase F (New England Biolabs) and incubated at 37 °C with agitation for 16 h. The negative control was incubated in the same buffer without PNGase F enzyme. The cell wall proteins were pelleted at 18000 rcf and washed 5 times in 50 mM Tris-HCl pH 8.0. The final cell wall pellet, equivalent to 12.5 mL of cell culture, was resuspended in 50 mM NH_4HCO_3, and proteins were digested with either trypsin 4 μg/mL (Sigma) or AspN 1 μg/mL (Promega) at 37 °C for 16 h with agitation.

Mass Spectrometry and Data Analysis

Peptides were desalted using C18 ZipTips (Millipore) and analyzed by LC−ESI−MS/MS using a Prominence nanoLC system (Shimadzu) and TripleTof 5600 mass spectrometer with a Nanospray III interface (AB SCIEX). Approximately 2 μg of peptides was desalted on an Agilent C18 trap (300 Å pore size, 5 μm particle size, 0.3 mm i.d. × 5 mm) at a flow rate of 30 μL/min for 3 min and then separated on a Vydac EVEREST reversed-phase C18 HPLC column (300 Å pore size, 5 μm particle size, 150 μm i.d. × 150 mm) at a flow rate of 1 μL/min. Peptides were separated with a gradient of 10−60% buffer B over 45 min, with buffer A (1% acetonitrile and 0.1% formic acid) and buffer B (80% acetonitrile with 0.1% formic acid). Gas and voltage setting were adjusted as required. An MS TOF scan from m/z of 350−1800 was performed for 0.5 s followed by information-dependent acquisition of MS/MS with automated CE selection of the top 20 peptides from m/z of 40−1800 for 0.05 s per spectrum. Peptides were identified using ProteinPilot (AB SCIEX), searching the LudwigNR database (downloaded from http://apcf.edu.au as at 27 January 2012; 16,818,973 sequences; 5,891,363,821 residues) with standard settings: Sample type, identification; Cysteine alkylation, acrylamide; Instrument, TripleTof 5600; Species, *S. cerevisiae* with common contaminants; ID focus, biological modifications; Enzyme, trypsin or AspN; Search effort, thorough ID. False discovery rate analysis using ProteinPilot was performed on all searches. Peptides identified with greater than 99% confidence and with a local false discovery rate of less than 1% were included for further analysis, and MS/MS fragmentation spectra were manually inspected. Extracted ion chromatograms were obtained using PeakView 1.1. Relative quantification of site-specific glycosylation occupancy was determined for each deamidated (deglycosylated)/unmodified (non-glycosylated) tryptic peptide pair without considering variable response factors of the unmodified and deamidated peptides. Occupancy was defined as the ratio of the extracted ion intensities of the deamidated peptide ion to the sum of the intensities of the deamidated and unmodified peptide ions.[17]

RESULTS

Identification and Relative Quantification of N-Glycosylation Site Occupancy in Cell Wall Proteins from Wild Type and Δ*alg3* Yeast Cells Using PNGase F Treatment Followed by Trypsin Digestion

We asked whether glycosylation occupancy in the yeast Δ*alg3* strain showed site-specific alterations. To efficiently detect multiple glycosylation sites, enrichment of glycoproteins is required.[17,28] Enrichment of yeast cell wall proteins ensures detection of both glycosylated and non-glycosylated forms of proteins since the covalent linkage of the proteins to the yeast polysaccharide cell wall matrix does not rely directly on N-glycans. Previous methods for analysis of site-specific N-glycosylation occupancy in yeast have used glycan release with Endo H.[17,18,23,24] However, the glycan structures on mature glycoproteins in the yeast Δ*alg3* mutant are altered, with the mutant glycan structures engendering the glycoproteins resistant to deglycosylation with Endo H.[29] We therefore deglycosylated cell wall enriched glycoproteins from Δ*alg3* and

dx.doi.org/10.1021/pr300599f | J. Proteome Res. 2012, 11, 5376−5383

Table 1. Peptides Containing N-Glycosylation Sequons Identified by Mass Spectrometry[a]

protein name	peptide position	sequence	Asn	m/z	z	Δmass
Gas1p	35−47	FFYSN**N**GSQFYIR	40	548.5878	3	−0.0001
Gas1p	35−47	FFYSNNGSQFYIR	40	548.2642	3	0.0132
Gas1p	48−77	GVAYQADTA**N**ETSGSTVNDPLANYESCSR	57	1031.453	3	0.0056
Gas1p	48−77	GVAYQADTANETSGSTVNDPLANYESCSR	57	1031.123	3	−0.0014
Gas1p	91−105	VYAI**N**TTLDHSECMK	95	599.6099	3	−0.0042
Gas1p	253−271	**N**LSIPVFFSEYGCNEVTPR	253	1122.536	2	0.0008
Gas5p	50−73	GVDYQPGGSS**N**LTDPLADASVCDR	60	837.0438	3	0.0022
Gas5p	50−73	GVDYQPGGSSNLTDPLADASVCDR	60	836.7169	3	0.0055
Pst1p	51−65	CDTLVG**N**LTIGGGLK	57	766.9019	2	−0.0025
Pst1p	51−65	CDTLVGNLTIGGGLK	57	766.4099	2	−0.0023
Ecm33p	318−335	GGANFDSSSS**N**FSCNALK	328	939.4022	2	0.0002
Ecm33p	318−335	GGANFDSSSSNFSCNALK	328	938.909	2	−0.0023
Pry3p	100−120	Y**N**YSNPGFSESTGHFTQVVWK	101	817.042	3	0.0019
Pry3p	100−120	YNYSNPGFSESTGHFTQVVWK	101	816.7147	3	0.0039
Yjl171 cP	219−240	**N**SSSIGYYDLPAIWLLNDHIAR	219	840.4252	3	0.0044
Yjl171 cP	219−240	NSSSIGYYDLPAIWLLNDHIAR	219	840.0991	3	0.0102
Plb2p	47−60	**N**ASGLSTAETDWLK	47	747.3575	2	−0.0042
Plb2p	47−60	NASGLSTAETDWLK	47	746.8659	2	−0.0032
Plb2p	185−200	SIVNPGGS**N**LTYTIER	193	574.6266	3	−0.0051
Plb2p	362−382	YVNNLSQDDDDIAIYAANPFK	365	796.0468	3	0.0061
Plb2p	491−509	**N**LTDLEYIPPLVVYIPNTK	491	1102.099	2	0.0019
Gas3p	268−288	L**N**STFEDAVIPLIFSEYGCNK	269	811.3958	3	0.0039
Gas3p	268−288	LNSTFEDAVIPLIFSEYGCNK	269	811.0667	3	0.0004
Crh1p	175−186	FH**N**YTLDWAMDK	177	771.3389	2	−0.0023
Crh1p	175−186	FHNYTLDWAMDK	177	770.8482	2	0.0002
Crh2p	23−59	ATFC**N**ATQACPEDKPCCSQYGECGTGQYCLNNCDVR	28	1098.453	4	0.0215
Crh2p	310−326	**N**GTSAYVYTSSSEFLAK	310	609.288	3	0.0003
Tos1p	412−421	AAVIFNSSDK	417	526.271	2	−0.0071
Yil169 cP	508−532	GEGVLAVDPTETNATPIPVVGYTGK	520	829.1041	3	0.0157

[a]Proteins covalently linked to the yeast polysaccharide cell wall were prepared, glycans were released by peptide-N-glycosidase F, and proteins were digested with trypsin, detected by LC−ESI−MS/MS, and identified by ProteinPilot. Deamidated asparagines are bold. Unglycosylated sequons are underlined. All methionines are oxidized and all cysteines are alkylated. All peptides listed were identified with >99% confidence.

wild type yeast strains with the general deglycosylating enzyme PNGase F, leaving previously glycosylated asparagines "tagged" with a deamidation ($m/z \approx 1$ Da). Deglycosylated covalently attached cell wall proteins were digested with trypsin, detected by LC−ESI−MS/MS, and identified with ProteinPilot (AB SCIEX). We identified peptides containing 19 NX(S/T) sequons as either previously glycosylated peptides modified with deamidation or non-glycosylated unmodified peptides. This gave a total of 29 peptides (>99% confidence) from 12 different proteins (Table 1, Supporting Information). We measured the relative glycosylation occupancy[17] at these 19 sites by measurement of deamidated-peptide/peptide pairs (Table 2 and Figure 1A and C). We found 10 N-glycosylation sites that showed decreased levels of glycosylation occupancy in Δ*alg3* compared to wild type. Ecm33p_328 showed the most severe decrease in glycosylation occupancy at 37% (Table 2 and Figure 1A−D). Some glycosylation sites showed only mild under-glycosylation, such as Crh1p_177 that showed 96% occupancy in the Δ*alg3* mutant (Table 2). We found six sites that were completely (100%) glycosylated in both Δ*alg3* and wild type cells, and three sites were found to be completely non-glycosylated in both wild type and mutant yeast cells (Table 2). Two glycosylation sites were mildly under-glycosylated in wild type cells, Ecm33p_328 (Table 2 and Figure 1 A and B) and Pry3p_101, and were both found to be further under-glycosylated in the Δ*alg3* strain (Table 2). In summary, we found 16 glycosylation sites that were occupied in wild type cells, 10 of which were under-glycosylated in the Δ*alg3* strain.

Spontaneous deamidation of NG sequences has been reported to occur during standard proteomic sample preparation.[30−32] It was therefore important to confirm that the deamidation events we detected were due to PNGase F catalyzed glycan release, rather than spontaneous chemical deamidation of non-glycosylated asparagines. Two of the detected glycosylation sites, Crh2p_310 and Gas1p_40, have sequons with the sequence NGS/T (Table 1) and would therefore be prone to spontaneous deamidation. We did not detect any deamidated peptide corresponding to Crh2p_310 in negative control samples not deglycosylated with PNGase F. In the negative control of Δ*alg3* cells we did detect a weak MS peak corresponding to ions of deamidated Gas1p_40 peptide, indicating that some portion of the unglycosylated form of this peptide underwent spontaneous deamidation during sample preparation (data not shown). However, no such peak was detected in the wild type negative control, indicating that this site is completely glycosylated in wild type and is partially under-glycosylated in Δ*alg3* cells. The intensity of the deamidated Gas1p_40 peptide in the negative control was very low and was neglected in calculation of relative glycosylation occupancy at this site, leading to a slight underestimation of the level of under-glycosylation at Gas1p_40 in the Δ*alg3* strain (Table 2). No detectable

dx.doi.org/10.1021/pr300599f | J. Proteome Res. 2012, 11, 5376−5383

Table 2. Site-Specific N-Glycosylation Occupancy in Δ*alg3* Yeast Strain[a]

	strain	
glycosylation site	wild type	Δ*alg3*
Gas1p_95	1.00 ± 0.00	1.00 ± 0.00
Gas1p_253	1.00 ± 0.00	1.00 ± 0.00
Plb2p_193	1.00 ± 0.00	1.00 ± 0.00
Plb2p_491	1.00 ± 0.00	1.00 ± 0.00
Crh2p_28	1.00 ± 0.00	1.00 ± 0.00
Crh2p_310	1.00 ± 0.00	1.00 ± 0.00
Gas5p_60	1.00 ± 0.00	0.97 ± 0.05
Crh1p_177	1.00 ± 0.00	0.96 ± 0.01
Gas1p_40	1.00 ± 0.00	0.94 ± 0.05
Gas3p_269	1.00 ± 0.00	0.92 ± 0.04
Pry3p_101	0.96 ± 0.02	0.85 ± 0.02
Pst1p_57	1.00 ± 0.00	0.76 ± 0.12
Plb2p_47	1.00 ± 0.00	0.66 ± 0.17
Gas1p_57	1.00 ± 0.00	0.57 ± 0.11
Yjl171cp_219	1.00 ± 0.00	0.51 ± 0.18
Ecm33_328	0.99 ± 0.01	0.37 ± 0.08
Plb2_365	0.00 ± 0.00	0.00 ± 0.00
Tos1p_417	0.00 ± 0.00	0.00 ± 0.00
Yil169cp_520	0.00 ± 0.00	0.00 ± 0.00

[a]The relative N-glycosylation occupancy at a given site was determined from the abundance of deamidated-modified and unmodified versions of the same sequon-containing peptide as measured by LC−ESI−MS/MS. Values are the mean of biological triplicates. Error is SD

spontaneous deamidation of asparagines in other sequons were detected in any of the negative controls.

Identification of N-Glycosylation Sites in Yeast Cell Wall Proteins with PNGase F Treatment Followed by AspN Digestion

Deglycosylation with PNGase F and digestion with trypsin allowed detection and relative quantification of site occupancy in Δ*alg3* cells. However, the mass shift upon deamidation is ∼1 Da, which in some cases can be difficult to differentiate from the normal isotopic distribution of non-deamidated peptides in the case that the deamidated and non-deamidated versions of that peptide are not separated by C18 LC prior to MS detection. This is in contrast to the ready differentiation of previously glycosylated asparagines tagged with a Δmass of 203.08 Da in the case of Endo H glycan release. This ambiguity has the potential to make measurement of site-specific N-glycan occupancy difficult. To overcome these difficulties and to further identify specific N-glycosylation sites showing under-glycosylation in Δ*alg3* yeast cells, we therefore treated the cell-wall glycoproteins with PNGase F followed by digestion with the protease AspN and detection with LC−ESI−MS/MS. Deamidation of a previously glycosylated asparagine converts the amino acid to an aspartic acid that will be recognized as a cleavage site for AspN. Consequently AspN will cleave at deamidated asparagines that were glycosylated prior to treatment with PNGase F. This will allow clear differentiation of previously glycosylated asparagines, in contrast to the difficulties of PNGase F/trypsin as described above. Using this approach we identified 49 peptides (>99% confidence) that had been cleaved at a deamidated asparagine in a sequon (glycosylated) or contained an unmodified sequon (non-glycosylated) (Table 3, Supporting Information). The 49 peptides represented 42 NX(S/T) sequons in 14 different proteins (Table 3); 37 of these sequons were identified as glycosylated. We found seven sequons that were present in unmodified peptides (Table 3). Three of these unmodified sequons (Ecm33p_328, Gas3p_350, and Crh2p_96) were found only in Δ*alg3* samples and were also detected as peptides cleaved at the corresponding deamidated asparagine in both wild type and Δ*alg3* cells, indicating partial glycosylation at these three sites in the Δ*alg3* mutant. Four sequons were identified exclusively as non-glycosylated. Two of these (Ecm33p_209 and Crh1p_177) were only detected from Δ*alg3* cells, indicating that these sites are under-glycosylated in the Δ*alg3* mutant strain (Table 3). Crh1p_177 was also identified in tryptic peptides both as a (deamidated) glycopeptide and a non-modified peptide (Table 1), showing some under-glycosylation at this site (Table 2). In total we identified five glycosylation sites that were under-glycosylated in the Δ*alg3* mutant strain using PNGase F followed by AspN treatment. In addition, we identified two sequons in non-modified peptides, Ccw14p_87 and Tos1p_417, present in both wild type and Δ*alg3* positive and negative controls (no PNGase F treatment), suggesting that these two sequons are not glycosylated in wild type or the Δ*alg3* mutant strain (Table 3). AspN cleavage due to spontaneous deamidation of an asparagine in an NGS sequon was detected at minor levels at Gas1p_40 in the Δ*alg3* negative control, indicating that a portion of this site that was not glycosylated in these cells underwent spontaneous chemical deamidation. Relative quantification of glycosylation site occupancy was not performed using AspN peptides due to the inability to directly compare the deamidated and unmodified forms of the same peptide.

Characteristics of Sequons That Showed Decreased Glycosylation Occupancy in Δ*alg3* Cells

N-Glycosylation is a general co-translational modification affecting many glycosylation sites on many different proteins within the cell. Only ∼70% of NX(S/T) sequons are glycosylated in proteins that are translocated into the ER.[17,18,23,24] However, the sequons that were normally glycosylated in wild type cells were not evenly affected in the Δ*alg3* mutant strain: some sites remained efficiently glycosylated, while others showed severe under-glycosylation. We compared the characteristics of glycosylation sites that were under-glycosylated and sites that showed no change in glycosylation occupancy in Δ*alg3* cells (Tables 2 and 4). We found that sequons that were efficiently glycosylated in Δ*alg3* cells were more likely to be located in loops of folded proteins (not in secondary structural elements) and contain threonine in the +2 position (Figure 2).

Identification of Novel N-Glycosylation Sites in Yeast Cell Wall Proteins

In *S. cerevisae* there are 196 N-glycosylation sequons that have been directly experimentally shown to be occupied (UniProt).[17,33] The method we describe here used cell wall enriched proteins that contain a limited number of proteins compared to a recently described method using filter-aided sample preparation (FASP) from whole cell extracts enriched for glycoproteins with ConA.[33] Nonetheless, using our approach we found in total 54 different occupied glycosylation sites (Tables 1 and 3), 26 of which have not been reported previously (Table 5), raising the total number of known occupied N-glycosylation sites in *S. cerevisae* to 222.

dx.doi.org/10.1021/pr300599f | J. Proteome Res. 2012, 11, 5376−5383

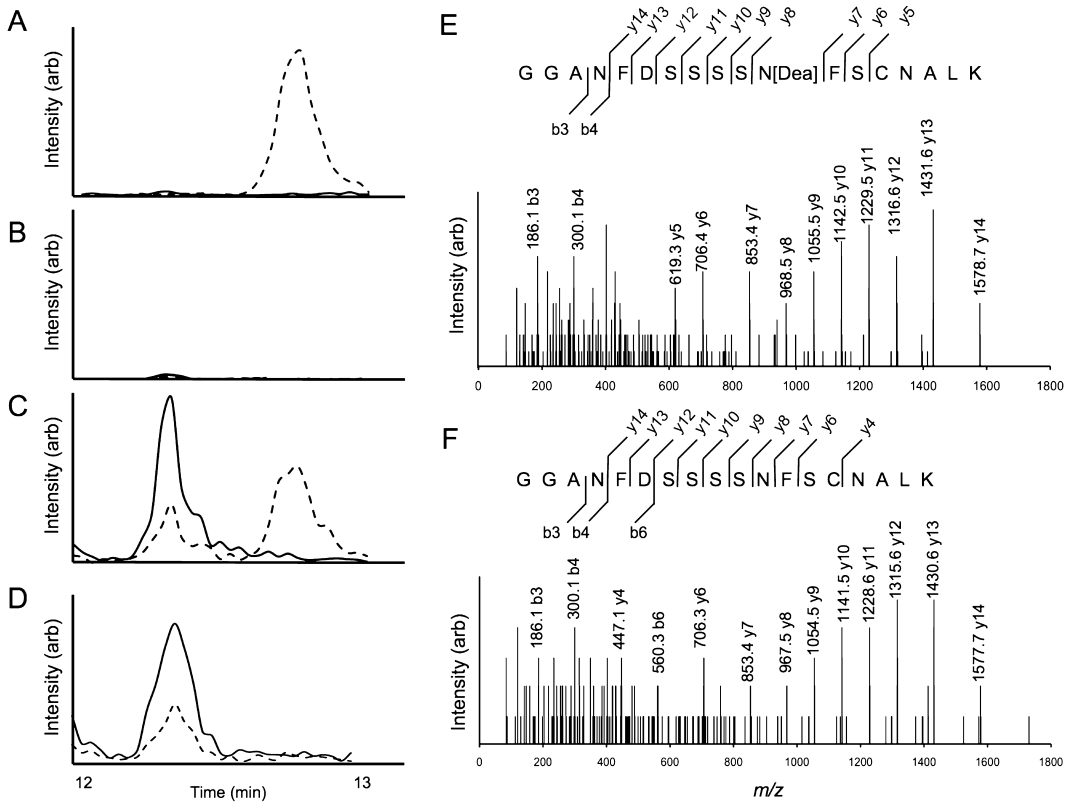


Figure 1. LC−ESI−MS/MS identification and relative quantification of glycosylation occupancy at Asn[328] of Ecm33p from wild type and Δ*alg3* cells. Extracted ion chromatograms of the $[M + 2H]^{2+}$ ions corresponding to deamidated peptide at an *m/z* of 939.40 (dashed line) and unmodified peptide at an *m/z* of 938.91 (solid line) from wild type (A, B) and Δ*alg3* (C, D). The glycoprotein -enriched cell wall proteins were treated with PNGase F followed by trypsin (A, C) or with trypsin (B, D). The ratio of the area under curve for the deamidated peptide to the total of deamidated and unmodified peptide gave a relative occupancy of 0.98 in wild type (A) and 0.37 in Δ*alg* (C) in this single measurement. The scale of the intensity is the same in panels A and B. MS/MS spectra and ProteinPilot matched fragmentation of deamidated peptide (E) and unmodified peptide (F) are shown.

■ DISCUSSION

We developed an LC−MS based analytical method to identify N-glycosylation sites and successfully determined the relative glycosylation occupancy at 19 sites using the yeast Δ*alg3* strain as a model. Independent of the yeast strain studied, deglycosylation with PNGase F followed by AspN digestion increased the number of glycosylation sites that could be detected in a complex protein sample at one time, when compared with trypsin digest. It appears that proteins with multiple N-glycosylation sites become more accessible to digest with AspN and MS analysis after treatment with PNGase F compared with Endo H. Of the 26 new sites we found in this study, 22 were identified using the combination of PNGase F and AspN. In total we could detect 54 different occupied glycosylation sites. The approach we describe here is therefore complementary to standard techniques and increases the number of N-glycosylation sites that can be detected at the same time on many different proteins.

While most glycosylation sequons were efficiently glycosylated in wild type cells, the altered glycan structure in the Δ*alg3* strain reduces the general glycosylation activity of OTase, leading to under-glycosylation.[26,29] However, glycosylation sites were not uniformly affected in Δ*alg3* cells, with the subset of sites with lower affinity for OTase most strongly affected. We found 13 specific glycosylation sites that were under-glycosylated in the Δ*alg3* strain compared to wild type cells (Table 4). The features of these glycosylation sites suggest that they are the subset of normally modified sites that are more difficult for OTase to glycosylate. For instance, these under-glycosylated sites are enriched in serine at the +2 position (Figure 2A). NxS sequons are glycosylated ∼40 times less efficiently than NxT sequons,[34] because of specific recognition of threonine at the +2 position by the peptide acceptor binding site of OTase.[35] Under-glycosylated sites in the Δ*alg3* strain were also likely to be present in secondary structural elements (helices or sheets), whereas efficiently glycosylated sites were more likely present in flexible loops (Figure 2B). Protein folding competes with glycosylation by OTase because the peptide acceptor binding site of OTase requires a flexible stretch of peptide for tight binding.[35] As secondary structure forms early in protein folding, glycosylation sites in secondary structural elements will have reduced affinity to OTase. The subset of glycosylation sites both present in flexible loops and with threonine at the +2 position were efficiently glycosylated (Figure 2C), suggesting a combination of these factors influences peptide acceptor binding to OTase and hence glycosylation efficiency.

Incorporation of either of the homologous proteins Ost3p or Ost6p into OTase results in enzyme isoforms with different protein substrate specificities at the level of individual glycosylation sites.[17] Six of the 13 sites that show under-glycosylation in Δ*alg3* cells have been shown to be under-glycosylated in yeast cells with double knockout of these homologous OTase subunits (Δ*ost3*/Δ*ost6*).[17] All six of these

dx.doi.org/10.1021/pr300599f | J. Proteome Res. 2012, 11, 5376−5383

Table 3. Peptides Containing N-Glycosylation Sequons Identified by Mass Spectrometry[a]

protein name	peptide position	peptide sequence	Asn	*m/z*	*z*	Δmass
Gas1p	23−39	D*DVPAIEVVGNKFFYSN.NGS	40	957.4678	2	0.0003
Gas1p	40−53	NGSQFYIRGVAYQA	40	525.5922	3	0.0021
Gas1p	77−94	DIPYLKKLNTNVIRVYAI.NTT	95	711.7537	3	0.0035
Gas5p	150−165	DNVLGFFAGNEVINSV.NTT	166	847.9219	2	−0.0020
Gas5p	166−179	NTTNTATYVKAVVR	166	769.9114	2	−0.0017
Gas5p	319−343	DFENLKNEYSKVSNPEGNGGYSTSN.NYS	344	917.4109	3	0.0007
Ecm33p	45−55	DKISGCSTIVG.NLT	56	575.7883	2	−0.0080
Ecm33p	82−91	NSSSLSSFSA	82	494.2102	2	−0.0135
Ecm33p	202−213	DNLVWANNITLR#	209	714.8843	2	0.0020
Ecm33p	214−226	DVNSISFGSLQTV.NAS	227	683.8457	2	−0.0007
Ecm33p	241−257	NLTQLSKVGQSLSIVSN	241	894.9885	2	−0.0006
Ecm33p	258−266	DELSKAAFS.NLT	267	484.2396	2	−0.0012
Ecm33p	279−286	NNTQLKVI	279	465.7654	2	−0.0020
Ecm33p	287−303	DGFNKVQTVGGAIEVTG.NFS	304	846.4332	2	−0.0007
Ecm33p	323−345	DSSSSNFSCNALKKLQSNGAIQG#	328	809.7242	3	0.0011
Ecm33p	328−345	NFSCNALKKLQSNGAIQG	328	655.6658	3	−0.0031
Pry3p	91−100	DAWYGEISKY.NYS	101	616.2849	2	−0.0005
Yjl171cP	286−299	DLGTGIQSYGYITR.NTT	300	772.392	2	0.0017
Ccw14p	73−91	DAAYSAFKSSCSEQNASLG	87	669.6309	3	0.0022
Plb1p	81−91	DTSLLSTLFGS.NSS	92	570.7927	2	−0.0001
Plb1p	265−276	DGRYPGTTVINL.NAT	277	653.3411	2	−0.0049
Plb1p	492−512	DLEYIPPLIVYIPNSRHSFNG.NQS	513	815.4273	3	0.0064
Plb2p	83−93	DTSLLSTLFSS.NSS	94	585.7971	2	−0.0019
Plb2p	162−170	NWTSVQEIV	162	538.7665	2	−0.0001
Plb2p	493−514	DLEYIPPLVVYIPNTKHSFNG.NQS	515	806.0919	3	0.0063
Plb2p	567−579	NATLPPECTKCFA	567	769.3527	2	−0.0045
Plb2p	615−629	DGIPITALLGSSTSG.NTT	630	694.8669	2	−0.0003
Gas3p	201−211	NRSIPVGYSAA	201	568.2903	2	−0.0009
Gas3p	191−200	DMKQYISKHA.NRS	201	610.8044	2	−0.0078
Gas3p	339−349	D*DFVNLESQLK.NVS	350	654.3276	2	0.0001
Gas3p	350−363	NVSLPTTKESEISS	350	746.8725	2	0.0000
Gas3p	339−363	D*DFVNLESQLKNVSLPTTKESEISS#	350	927.4709	3	0.0145
Gas3p	370−384	DNSAITNIYSGFGTN.NFT	385	787.36	2	0.0000
Gas3p	414−421	DYAVPTTF.NYT	422	457.219	2	0.0005
Crh1p	117−129	NGTGIVSSFYLQS	117	687.3319	2	−0.0017
Crh1p	173−180	DKFHNYTL#	177	519.2554	2	−0.0016
Crh1p	192−200	DGESVRVLS.NTS	201	481.2486	2	−0.0051
Crh2p	82−95	DYSSKLGNANTFLG.NVS	96	743.8612	2	−0.0021
Crh2p	82−100	DYSSKLGNANTFLGNVSEA#	96	993.9658	2	−0.0160
Crh2p	175−189	DLETAQTNFYWESVL.NYT	190	908.4202	2	−0.0105
Crh2p	223−232	DGVVGRTLYK.NET	233	554.3093	2	−0.0044
Crh2p	237−252	NATTQKYQYPQTPSKV	237	927.9626	2	−0.0052
Crh2p	253−260	DISIWPGG.NST	261	422.7143	2	0.0013
Crh2p	282−296	DISNPGYYYAIVNEV.NIT	297	858.9086	2	−0.0016
Crh2p	310−326	NGTSAYVYTSSSEFLAK	310	913.4259	2	−0.0045
Tos1p	222−235	DWSRGSYFVPGSTS.NCT	236	773.3513	2	−0.0016
Tos1p	407−419	DGTLKAAVIFNSS	417	661.8506	2	−0.0011
Pir3p	231−244	NSTLSMSLSKGILT	231	726.8838	2	−0.0011
Pst1p	215−227	DVHSVSFANLQKI.NSS	228	729.3972	2	0.0124

[a]Proteins covalently linked to the yeast polysaccharide cell wall were prepared, glycans were released by peptide-N-glycosidase F, and proteins were digested with AspN, detected by LC−ESI−MS/MS, and identified by ProteinPilot. Deamidated asparagines are bold. Unglycosylated sequons are underlined. *, missed cleavage; #, peptides containing unglycosylated sequons identified in Δ*alg3* only. All methionines are oxidized and all cysteines are alkylated. All peptides listed were identified with >99% confidence.

sites were shown to be rescued by overexpression of Ost3p, although two sites could be rescued by overexpression of either Ost3p or Ost6p.[17] Furthermore, one glycosylation site, Gas1p_253, was shown to require Ost6p for complete glycosylation occupancy.[17] Interestingly, we found that glycosylation occupancy of Gas1p_253 was not affected in Δ*alg3* cells (Table 2), which express both Ost3p- and Ost6p-containing OTase isoforms. This emphasizes the importance of the protein substrate specificities of OTase isoforms in defining the characteristics of an efficiently glycosylated asparagine.

Our results may be complicated by the ability of many proteins to fold correctly even without glycosylation at many

dx.doi.org/10.1021/pr300599f | J. Proteome Res. 2012, 11, 5376−5383

Table 4. Summary of N-Glycosylation Sites That Are Under-glycosylated in Δ*alg3*

glycosylation site	phenotype[a]	identified with
Gas1p_40	mild	trypsin
Gas1p_57	strong	trypsin
Gas5p_60	mild	trypsin
Pst1p_57	medium	trypsin
Ecm33_209	n.a.	AspN
Ecm33_328	strong	trypsin/AspN
Pry3p_101	medium	trypsin
Yjl171cp_219	strong	trypsin
Plb2_47	strong	trypsin
Gas3p_269	mild	trypsin
Gas3p_350	n.a.	AspN
Crh1p_177	mild	trypsin/AspN
Crh2p_96	n.a.	AspN

[a]n.a.: not applicable.

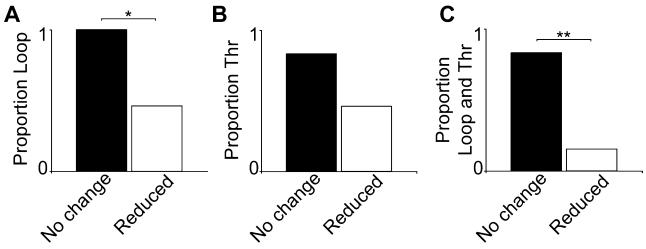


Figure 2. Characteristics of glycosylation sites that are under-glycosylated in Δ*alg3* cells. Proportion of glycosylation sites that are present in (A) loops, (B) NxT sequons, or (C) NxT sequons in loops that show no change in occupancy (black) or that are under-glycosylated (white) in Δ*alg3* cells compared with wild type cells. * $p < 0.05$, ** $p < 0.01$, Fisher's exact test.

Table 5. Novel N-Glycosylation Sites Detected Using PNGase F

Uniprot ID	protein name	occupied N-glycosylation sites
Q08193	Gas5p	Asn 166
P38248	Ecm33p	Asn 56, 82, 227, 241, 304
P47033	Pry3p	Asn 101
P46992	Yjl171cP	Asn 219, 300
P39105	Plb1p	Asn 92, 277
Q03674	Plb2p	Asn 47, 94, 162, 515, 567
Q03655	Gas3p	Asn 269, 385, 422
P53301	Crh1p	Asn 201
P32623	Crh2p	Asn 190, 237, 261, 297
Q03180	Pir3p	Asn 231
Q03180	Pst1p	Asn 57

sites, as long as a certain critical level of glycosylation is present. This will mean that if glycosylation is strictly required at a particular asparagine for correct protein folding, then that asparagine will appear to always be glycosylated, even though most of the nascent polypeptide that is not modified will be degraded by the quality control systems of the ER.[2] In addition, glycosylation efficiency may depend on factors such as local folding of the target protein close to the sequon and the protein substrate specificities of the particular OTase isoforms present.[17,36] However, in combination this and previous studies have detected almost every sequon in both glycosylated and non-glycosylated forms, suggesting that these effects are relatively minor. Therefore, it is reasonable to imply that sites that are easier to glycosylate would be expected to show full occupancy and sites that are more difficult to glycosylate would be found under-glycosylated.

CONCLUSION

Mutations in the human ALG3 gene cause a rare form of CDG (CDG-Id).[37] Reduced OTase activity due to the altered structure of the lipid-linked glycan substrate in this disorder is likely to affect N-glycosylation occupancy in a site-specific manner, similar to the patterns we observe here in a *S. cerevisiae* model system. We predict that glycosylation sites in NxS sequons and those located in secondary structural elements will be the most likely to show reduced glycosylation in CDG-Id. Many different diseases and health conditions in addition to the CDG spectrum are caused by or associated with defects in glycosylation and consequently decreased levels of N-glycosylation occupancy at many different glycosylation sites in many different proteins. In conjunction with appropriate glycoprotein or glycopeptide enrichment strategies, the method we describe here could potentially be further developed to study human N-glycosylation site occupancy using clinical tissue, serum, or saliva protein samples, leading to discovery of new biomarkers for CDG and other disorders and improved diagnostics.

ASSOCIATED CONTENT

Supporting Information

This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +61 7 3365 4875. Fax: +61 7 3365 4273. E-mail: b.schulz@uq.edu.au.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the support of NHMRC project grant 631615 and CDF APP1031542

REFERENCES

(1) Kelleher, D. J.; Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006**, *16* (4), 47R−62R.

(2) Schulz, B. L., Beyond the sequon: sites of N-glycosylation. In *Glycosylation*; Petrescu, S., Ed.; InTech: New York, **2012**; available from: http://www.intechopen.com/books/glycosylation/beyond-the-sequon-sites-of-n-glycosylation.

(3) Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **2004**, *73*, 1019−49.

(4) Chen, W.; Helenius, A. Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. *Mol. Biol. Cell* **2000**, *11* (2), 765−72.

(5) Whitley, P.; Nilsson, I. M.; von Heijne, G. A nascent secretory protein may traverse the ribosome/ER translocase complex as an extended chain. *J. Biol. Chem.* **1996**, *271*, 6241−44.

(6) Haeuptle, M. A.; Hennet, T. Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum. Mutat.* **2009**, *30* (12), 1628−41.

(7) Janik, M. E.; Litynska, A.; Vereecken, P. Cell migration-the role of integrin glycosylation. *Biochim. Biophys. Acta* **2010**, *1800* (6), 545−55.

(8) Sato, S.; St-Pierre, C.; Bhaumik, P.; Nieminen, J. Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). *Immunol. Rev.* **2009**, *230* (1), 172−87.
(9) Bova, G. S.; Carter, B. S.; Bussemakers, M. J. G.; Emi, M.; Fujiwara, Y.; Kyprianou, N.; Jacobs, S. C.; Robinson, J. C.; Epstein, J. I.; Walsh, P. C.; Isaacs, W. B. Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. *Cancer Res.* **1993**, *53* (17), 3869−73.
(10) Honma, K.; Iwao-Koizumi, K.; Takeshita, F.; Yamamoto, Y.; Yoshida, T.; Nishio, K.; Nagahara, S.; Kato, K.; Ochiya, T. RPN2 gene confers docetaxel resistance in breast cancer. *Nat. Med.* **2008**, *14* (9), 939−48.
(11) Pils, D.; Horak, P.; Gleiss, A.; Sax, C.; Fabjani, G.; Moebus, V. J.; Zielinski, C.; Reinthaller, A.; Zeillinger, R.; Krainer, M. Five genes from chromosomal band 8p22 are significantly down-regulated in ovarian carcinoma. *Cancer* **2005**, *104* (11), 2417−29.
(12) Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1999**, *1473* (1), 4−8.
(13) Hülsmeier, A. J.; Paesold-Burda, P.; Hennet, T. N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography mass spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (12), 2132−8.
(14) Garshasbi, M.; Hadavi, V.; Habibi, H.; Kahrizi, K.; Kariminejad, R.; Behjati, F.; Tzschach, A.; Najmabadi, H.; Ropers, H. H.; Kuss, A. W. A defect in the TUSC3 gene is associated with autosomal recessive mental retardation. *Am. J. Hum. Genet.* **2008**, *82* (5), 1158−64.
(15) Mohorko, E.; Glockshuber, R.; Aebi, M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.* **2011**, *34* (4), 869−78.
(16) Molinari, F.; Foulquier, F.; Tarpey, P. S.; Morelle, W.; Boissel, S.; Teague, J.; Edkins, S.; Futreal, P. A.; Stratton, M. R.; Turner, G.; Matthijs, G.; Gecz, J.; Munnich, A.; Colleaux, L. Oligosaccharyltransferase-subunit mutations in non-syndromic mental retardation. *Am. J. Hum. Genet.* **2008**, *82* (5), 1150−7.
(17) Schulz, B. L.; Aebi, M. Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-glycosylation efficiency. *Mol. Cell Proteomics* **2009**, *8* (2), 357−64.
(18) Schulz, B. L.; Stirnimann, C. U.; Grimshaw, J. P. A.; Brozzo, M. S.; Fritsch, F.; Mohorko, E.; Capitani, G.; Glockshuber, R.; Grütter, M. G.; Aebi, M. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (27), 11061−6.
(19) Burda, P.; Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta* **1999**, *1426* (2), 239−57.
(20) Runge, K. W.; Robbins, P. W. A new yeast mutation in the glucosylation steps of the asparagine-linked glycosylation pathway. Formation of a novel asparagine-linked oligosaccharide containing two glucose residues. *J. Biol. Chem.* **1986**, *261* (33), 15582−90.
(21) Henquet, M.; Eigenhuijsen, J.; Spiegel, H.; Schreuder, M.; van Duijn, E.; Cordewener, J.; Depicker, A.; van der Krol, A.; Bosch, D. Characterization of the single-chain Fv-Fc antibody MBP10 produced in Arabidopsis alg3 mutant seeds. *Transgenic Res.* **2011**, *20* (5), 1033−42.
(22) Henquet, M.; Lehle, L.; Schreuder, M.; Rouwendal, G.; Molthoff, J.; Helsper, J.; van der Krol, A.; Bosch, D. Identification of the gene encoding the alpha1,3-mannosyltransferase (ALG3) in Arabidopsis and characterization of downstream n-glycan processing. *Plant Cell* **2008**, *20* (6), 1652−64.
(23) Izquiedro, L.; Schulz, B. L.; Rodrigues, J. A.; Güther, M. L.; Procter, J. B.; Barton, G. J.; Aebi, M.; Ferguson, M. A. Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J.* **2009**, *28* (17), 2650−61.
(24) Nasab, F.; Schulz, B.; Gamarro, F.; Parodi, A.; Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol. Biol. Cell* **2008**, *19* (9), 3758−68.
(25) Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat. Biotechnol.* **2003**, *21* (6), 660−6.
(26) Huffaker, T. C.; Robbins, P. W. Yeast mutants deficient in protein glycosylation. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80* (24), 7466−70.
(27) Yin, Q. Y.; de Groot, P. W.; Dekker, H. L.; de Jong, L.; Klis, F. M.; de Koster, C. G. Comprehensive proteomic analysis of Saccharomyces cerevisiae cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. *J. Biol. Chem.* **2005**, *280* (21), 20894−901.
(28) Schulz, B. L.; White, J. C.; Punyadeera, C. Saliva proteome research: current status and future outlook. *Crit. Rev. Biotechnol.* **2012**, DOI: 10.3109/07388551.2012.687361.
(29) Aebi, M.; Gassenhuber, J.; Domdey, H.; te Heesen, S. Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae. *Glycobiology* **1996**, *6* (4), 439−44.
(30) Krokhin, O. V.; Antonovici, M.; Ens, W.; Wildins, J. A.; Standing, K. G. Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal. Chem.* **2006**, *78* (18), 6645−50.
(31) Robinson, N. E.; Robinson, Z. W.; Robinson, B. R.; Robinson, A. L.; Robinson, J. A.; Robinson, M. L.; Robinson, A. B. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J. Pept. Res.* **2004**, *63* (5), 426−36.
(32) Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R. Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* **2012**, *11* (3), 1949−57.
(33) Breidenbach, M. A.; Palaniappan, K. K.; Pitcher, A. A.; Bertozzi, C. R. Mapping yeast N-glycosites with isotopically recoded glycans. *Mol. Cell Proteomics* **2012**, *11* (6), M111.015339.
(34) Kasturi, L.; Eshleman, J. R.; Wunner, W. H.; Shakin-Eshleman, S. H. The hydroxy amino acid in an Asn-X-Ser/Thr sequon can influence N-linked core glycosylation efficiency and the level of expression of a cell surface glycoprotein. *J. Biol. Chem.* **1995**, *270* (24), 14756−61.
(35) Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011**, *474* (7351), 350−5.
(36) Jamaluddin, M. F. B.; Bailey, U. M.; Tan, N. Y. J.; Stark, A. P.; Schulz, B. L. Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Sci.* **2011**, *20* (5), 849−55.
(37) Sun, L.; Eklund, E. A.; Chung, W. K.; Wang, C.; Cohen, J.; Freeze, H. H. Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. *J. Clin. Endocrinol. Metab.* **2005**, *90* (7), 4371−5.

Chapter 5

Sequence-based protein stabilization in the absence of glycosylation

5.1 Introduction to this publication

This chapter was published in *Nature Communications* as an original article. In this section, I briefly describe my role and contribution towards this successful work. I used a novel strategy for engineering glycoprotein stability to determine the detailed function of Ost3p and Ost6p *in vitro*. The ER lumenal domain of Ost3p has a single glycosylation sequon, while Ost6p has no glycosylation sequons and contains a cluster of acidic amino acids at the position corresponding to the Ost3p sequon. The ER lumenal domain of Ost6 is more easily expressed than Ost3 in *E.coli* and shows *in vitro* peptide binding activity when Ost3 does not. We showed that incorporation of targeted like-charged amino acids at *N*-glycosylation sequons in Ost3/6p increases *in vitro* stability and activity. Optimal stabilization through introducing a minimal number of local point mutations at the Ost3p glycosylation sequon allowed increased *in vitro* binding of peptides. These peptides were complementary to the characteristics of the peptide binding groove of Ost3, a key determinant of transient nascent polypeptide tethering by Ost3p, consistent with the function of Ost3p and Ost6p in modulating *N*-glycosylation substrate selection by OTase activity *in vivo*.

5.2 Declaration on authorship

In accordance with The University of Queensland, the methods, results and discussion are presented in the form of a published article in the peer-reviewed journal. This article was published in *Nature Communications*. The contributions of all authors are mentioned in Section 5.3.

5.3 Published peer-reviewed article

"Sequence-based protein stabilization in the absence of glycosylation"

Tan NY, Bailey UM, **Jamaluddin MF**, Mahmud SH, Raman SC, Schulz BL. Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun*. **2014**, 5: 3099.

Contributions:

This paper describes a study of how protein sequences can compensate for the absence of glycosylated sequons.

This work presented here was done under the supervision of Dr Benjamin Schulz.

Ms Nikki Tan Yi Jie., Mr Suresh Raman and Dr Benjamin Schulz performed sequence analyses. Ms Nikki Tan Yi Jie., Ms Siti Halimah Mahmud and Dr Benjamin Schulz characterized the proteins. Dr Maja Bailey performed the growth inhibition assay. I performed molecular biology (cloning), protein expression and peptide affinity chromatography assay. Dr Benjamin Schulz and I performed mass spectrometry. All authors contributed equally in writing manuscript. All authors read and approved the manuscript.

nature COMMUNICATIONS

ARTICLE

Received 14 Aug 2013 | Accepted 12 Dec 2013 | Published 17 Jan 2014

DOI: 10.1038/ncomms4099

Sequence-based protein stabilization in the absence of glycosylation

Nikki Y. Tan[1], Ulla-Maja Bailey[1], M Fairuz Jamaluddin[1], S Halimah Binte Mahmud[1], Suresh C. Raman[1] & Benjamin L. Schulz[1]

Asparagine-linked *N*-glycosylation is a common modification of proteins that promotes productive protein folding and increases protein stability. Although *N*-glycosylation is important for glycoprotein folding, the precise sites of glycosylation are often not conserved between protein homologues. Here we show that, in *Saccharomyces cerevisiae*, proteins upregulated during sporulation under nutrient deprivation have few *N*-glycosylation sequons and in their place tend to contain clusters of like-charged amino-acid residues. Incorporation of such sequences complements loss of *in vivo* protein function in the absence of glycosylation. Targeted point mutation to create such sequence stretches at glycosylation sequons in model glycoproteins increases *in vitro* protein stability and activity. A dependence on glycosylation for protein stability or activity can therefore be rescued with a small number of local point mutations, providing evolutionary flexibility in the precise location of *N*-glycans, allowing protein expression under nutrient-limiting conditions, and improving recombinant protein production.

[1] School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia. Correspondence and requests for materials should be addressed to B.L.S. (email: b.schulz@uq.edu.au).

The presence of *N*-glycans assists protein folding, increases protein solubility, reduces aggregation and limits proteolysis[1–5], while their specific structures affect glycoprotein function in development, cancer, inflammation and infection[3,6]. In eukaryotes, *N*-glycans are added to selected asparagine residues in nascent polypeptides co- and post-translocationally in the lumen of the endoplasmic reticulum (ER) by oligosaccharyltransferase (OTase)[7–10]. *N*-glycosylation by OTase occurs before substrate protein folding, as only flexible polypeptide substrate can bind at its active site[11]. The presence of Ser or Thr at the +2 position markedly increases glycosylation efficiency of asparagines, due to specific interactions with the peptide-binding site of OTase[11], and such sequences are termed glycosylation sequons (N-x-S/T; x≠P). Asparagines in secondary structural elements tend to be inefficiently glycosylated[12,13], as their reduced flexibility lowers their affinity to the OTase active site[11]. *N*-glycosylation therefore often occurs between secondary structural elements of folded proteins[14]. This allows *N*-glycans to optimally assist productive protein folding by their hydrophilic bulk lowering the entropy of the protein unfolded state and acting as nucleation sites for local folding of nascent glycopolypeptide chains[2]. *N*-glycans can also increase protein stability by making direct contacts with amino acids in the folded protein[15,16], increasing overall protein solubility[17], shielding exposed hydrophobic surfaces[18] and limiting proteolysis[19]. In the ER, *N*-glycans act as signals to locally recruit ER chaperones to assist in glycoprotein oxidative folding and are important in glycoprotein quality control and degradation[2,20–22]. Although *N*-glycans are important for protein folding and function, the precise sites of glycosylation within a protein are not generally conserved in homologues.

Here, we study the extent to which peculiar amino-acid sequences can functionally replace *N*-glycosylation sites. We show that clusters of like-charged amino acids in place of *N*-glycosylation sequons can impart local protein stability and improve protein function in the absence of glycosylation.

Results

Sequence analysis of yeast cell wall *N*-glycoproteins. In the absence of a nitrogen source and in the presence of a non-fermentable carbon source, bakers' yeast *S. cerevisiae* ceases vegetative growth and enters sporulation[23]. This results in large changes to many aspects of yeast biology including remodelling of the cell wall. Key enzymes involved in this process are the *GAS* family of glucanases/transglycosidases. Gas1p is a major cell-wall protein during vegetative growth while its functionally equivalent paralogue Gas2p is predominantly expressed during sporulation[24,25]. Gas1p and Gas2p share 45% sequence identity, conserved secondary structure and identical disulphide bond connectivity[26,27]. However, Gas1p has 10 *N*-glycosylation sequons while Gas2p has only 1. This prompted us to test if other sporulation-induced proteins also had few glycosylation sequons. This analysis showed that glycoproteins upregulated in sporulation did have fewer glycosylation sequons than their paralogues expressed during vegetative growth (Fig. 1a). The nutrient limitations that induce sporulation may provide a selective pressure for reducing flux through the energetically expensive *N*-glycosylation pathway, resulting in loss of *N*-glycosylation sequons from sporulation-specific proteins.

Given the importance of *N*-glycosylation for protein folding, we investigated if the equivalent positions of *N*-glycans in paralogues that lacked a sequon at that particular site had compensatory amino-acid sequences. For this, we used a set of experimentally validated yeast glycosylation sites[13,28–30]. Sequence analysis of positions in paralogues corresponding to these sites showed that they had more acidic amino acids compared with both validated glycosylation sites and the average amino-acid composition of the whole protein set (Fig. 1b–d). This enrichment was only significant in the immediate vicinity of the residues corresponding to the location of the glycosylation sequon (Fig. 1e–h), although the high sequence diversity in this analysis precluded the identification of more specific patterns.

Functional complementation of glycosylation sites in yeast. To test if these clusters of acidic amino acids served to provide local intrinsic stability to the protein even in the absence of *N*-glycosylation, we expressed and purified the model glycoproteins yeast Gas1p and Gas2p as secreted MBP fusions in *E. coli* and *S. cerevisiae*. Given Gas1p has many more *N*-glycosylation sequons than Gas2p, we predicted that Gas2p would be more stable than Gas1p when expressed without *N*-glycosylation in *E. coli*. Circular dichroism (CD) spectroscopy of *E. coli*-expressed protein was consistent with MBP-Gas2 containing more helical secondary structure than MBP-Gas1, while the native glycosylated forms of these *S. cerevisiae*-expressed proteins showed equivalent CD spectra (Fig. 2a,b). MS analysis of the disulphide bond connectivity by non-reducing tryptic peptide mapping showed that both *E. coli*-expressed MBP-Gas1 and MBP-Gas2 had incorrect connectivity compared with native protein expressed in yeast (Supplementary Tables 1 and 2)[26,27]. To enable comparison with native glycosylated MBP-Gas1 and MBP-Gas2, we created disulphide shuffled forms of these *S. cerevisiae*-expressed proteins. CD spectra of these disulphide shuffled forms of MBP-Gas1 and MBP-Gas2 were similar to their native forms (Fig. 2a,b). This is consistent with the increased stability and secondary structure in nonglycosylated MBP-Gas2 compared with nonglycosylated MBP-Gas1 being local effects that influence secondary structure formation. That is, the absence of *N*-glycosylation at the 10 glycosylation sites in Gas1p strongly affected local formation of secondary structural elements, whereas this affect was not as strong in Gas2p, which with only a single glycosylation site does not natively rely on *N*-glycosylation for stability to the extent of Gas1p.

We next tested if the Gas2p sequences corresponding to Gas1p glycosylation sites were functionally equivalent *in vivo*. Yeast cells with deletion of *GAS1* showed reduced growth compared with wild-type cells, which was complemented by overexpression of native Gas1p (Fig. 2c). Gas1p is glycosylated at Asn40 ($NN_{40}GS$)[13] and expression of variant Gas1-N40Q lacking this glycosylation site did not complement loss of *GAS1* (Fig. 2c). While Asn-Gln site-directed mutagenesis is routinely used to test the importance of particular glycosylation sites in glycoprotein folding and function, the conformational preferences of Gln are very different to those of Asn, while His is more similar to Asn[31]. Expression of variant Gas1-N40H only partially complemented loss of *GAS1* (Fig. 2c). This lack of complementation indicated that glycosylation at Asn40 was required for Gas1p function in the context of the native sequence. However, replacing this stretch of Gas1p with the corresponding sequence from Gas2p ($_{39}ESGE_{42}$; Supplementary Fig. 1a) restored growth, as did point mutation to insert proximal Asp residues (Fig. 2c), while retaining the conserved Gly41 (Supplementary Fig. 1b). The dependency of *in vivo* Gas1p function on glycosylation at Asn40 could therefore be rescued by a small number of compensatory point mutations introducing local clusters of negatively charged residues.

Stabilization of unglycosylated human interferon-β. Animal glycoproteins are not generally required to be expressed under such severe nutrient limitations as during yeast sporulation.

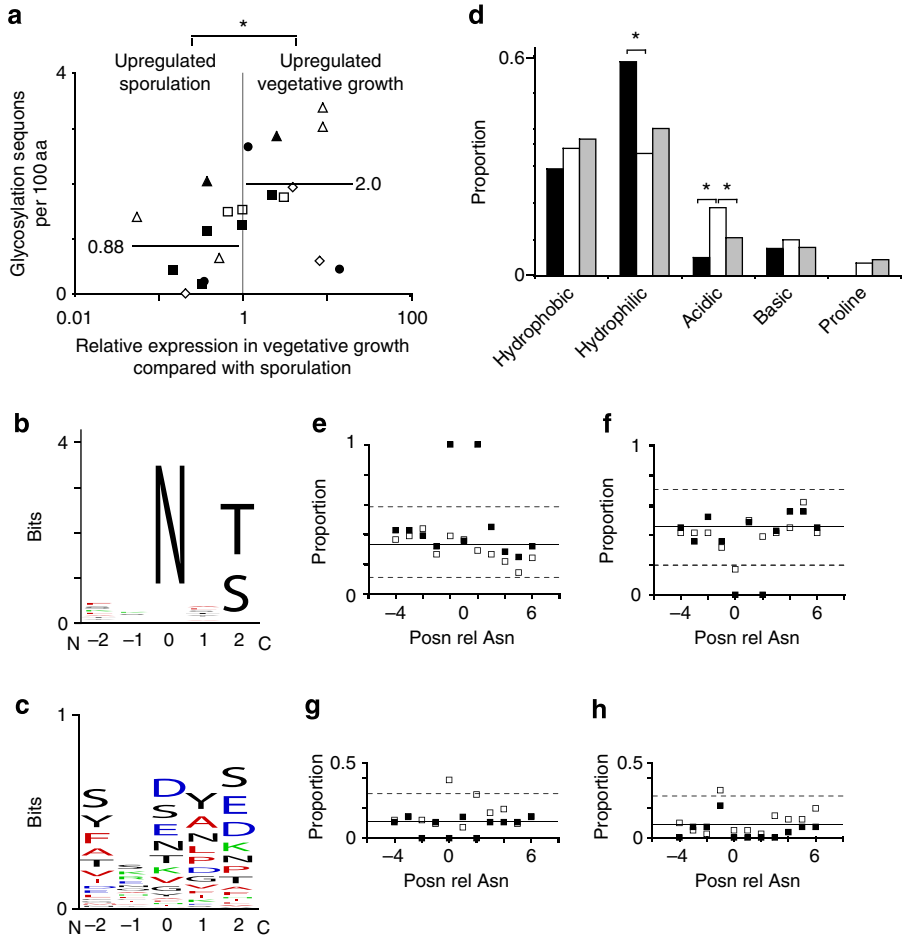


Figure 1 | Equivalent sequences to glycosylation sites in yeast cell-wall proteins. (**a**) Glycosylation sequon density is higher in glycoprotein paralogues upregulated in vegetative growth than in those upregulated in sporulation. Paralogues of black squares, Gas1p; white squares, Mkc7p; black circles, Exg1p; black triangles, Plb1p; white triangles, Ecm33p; and white diamonds, Crh1p. Horizontal bars show mean for each group. *$P = 0.01$, 2-sided Mann–Whitney test. (**b**) Weblogo representation of aligned experimentally validated glycosylation sites and (**c**) corresponding sequences in homologues lacking glycosylation sites. (**d**) Proportion of hydrophilic, hydrophobic, acidic and basic amino acids in validated sites (black), corresponding sequences lacking sequons (white) and in the entire protein sequences of all homologues analysed (grey). *$P < 0.005$, Fisher's exact test. Position specific proportion of (**e**) hydrophilic, (**f**) hydrophobic, (**g**) acidic and (**h**) basic amino acids for validated glycosylation sites in yeast cell-wall proteins (black squares) and corresponding sequences in paralogues lacking sites (white squares). Horizontal lines indicate mean values of the entire protein sequence of all homologues analysed (full) and the values at which $P = 0.001$ by Fisher's exact test compared with this mean for a given position (dashed).

However, human glycoproteins are often expressed for biotechnological purposes in bacteria that lack glycosylation systems. Such a protein is the human glycoprotein hormone interferon-β (IFNβ)[32]. The single *N*-glycosylation site in human IFNβ is not strictly conserved in orthologues and paralogues (Fig. 3a, Supplementary Fig. 2), and in these proteins is often replaced by clusters of acidic amino acids similar to those observed in place of *N*-glycosylation sites in yeast proteins (Fig. 1, Supplementary Fig. 1). To test if such a sequence would increase IFNβ stability in the absence of glycosylation, we expressed and purified IFNβ as wild type and glycosylation site variant (Asn101Asp) forms from *E. coli* and refolded these proteins *in vitro*. Guanidinium-induced unfolding equilibria showed that at physiologically relevant pH 7.5 this point mutation increased stability ($\Delta\Delta G^0$ $-9\,kJ\,mol^{-1}$, Fig. 3b). In contrast, at pH 5.0 the $\Delta\Delta G^0$ was negligible (Fig. 3c). We propose that this is consistent with stabilization of the $IFN\beta_{N101D}$ variant at pH 7.5 through reduced entropy of the unfolded state due to charge-charge repulsion between Asp101 and Glu102 favouring a local turn or extended structure similar to the structure of Asn101 and Glu102 in folded native IFNβ[18], or stabilization of the native state[33,34]. At pH 5 there is a lower probability that both these residues will be simultaneously deprotonated (negatively charged) in a given protein molecule, consistent with the lack of stabilization in the $IFN\beta_{N101D}$ variant. The overall higher stability for both protein forms at pH 5.0 than at pH 7.5 is likely due to increased distance from the predicted pI of IFNβ (~9.0). This confirmed that local clusters of negatively charged amino acids could increase protein stability in the absence of glycosylation in homologues of human IFNβ. To confirm that stable $IFN\beta_{N101D}$ retained native activity, we performed a cell culture growth inhibition assay (Fig. 3d). Stable $IFN\beta_{N101D}$ showed significantly higher specific activity than IFNβ wild type (Fig. 3d), correlating increased thermodynamic stability with improved protein function in this assay. This suggested that it is possible for IFN proteins to have either *N*-glycosylation sites or sequence stretches that can play a similar role in protein folding and stability.

Stabilization at an inefficiently glycosylated sequon. Asparagines in sequons generally show macroheterogeneity with not all of the protein pool glycosylated, and almost all sequons are

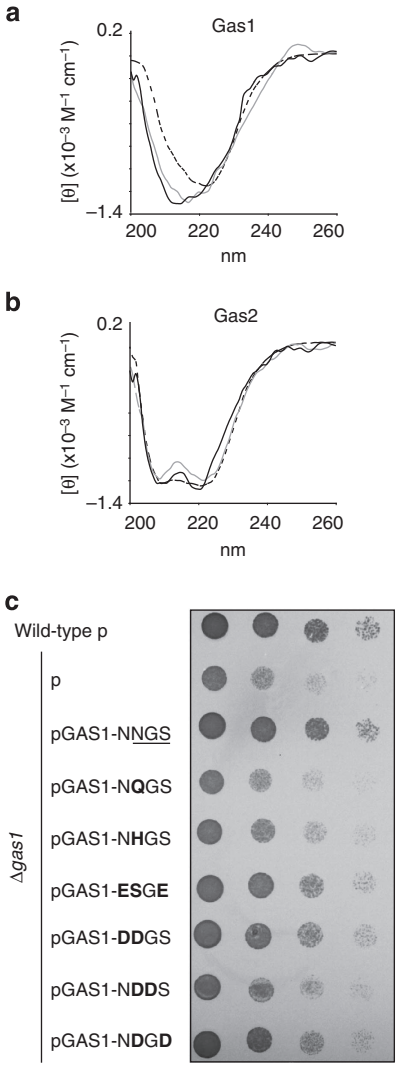


Figure 2 | Structure and function of Gas1p and Gas2p glycosylation site variants. (**a**) CD spectra of MBP-Gas1 and (**b**) MBP-Gas2 produced in *E. coli* (dashed black), or *S. cerevisiae* native (solid black) or disulphide-shuffled (solid gray). Mean residue ellipticity is shown. (**c**) Growth of yeast with wild-type Gas1p (wild type, pGAS1-NNGS) and Asn40 glycosylation site variants. The native Asn40 sequon is underlined. Point mutations in variants are bold.

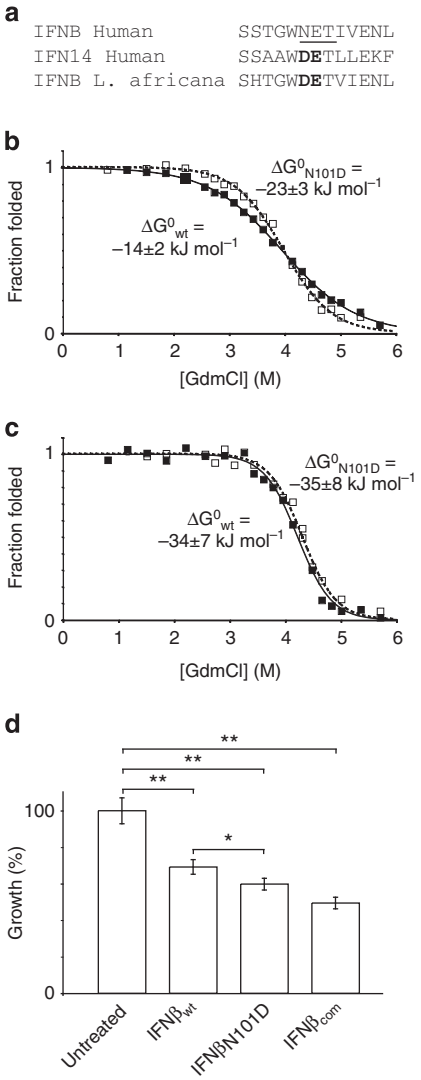


Figure 3 | Stability and activity of human IFNβ variants. (**a**) Alignment of human IFNβ and a selected orthologue and paralogue. The IFNβ glycosylation sequon is underlined and corresponding clusters of acidic amino acids are in bold. (**b**) Guanidinium induced equilibrium unfolding of IFNβ wild-type and $IFN\beta_{N101D}$ variant as measured by intrinsic tryptophan fluorescence and fitted to the two-state model of protein folding at pH 7.5 and at (**c**) pH 5.0. Solid line, IFNβ wild type; dashed line, $IFN\beta_{N101D}$. (**d**) Activity of $IFN\beta_{wt}$, $IFN\beta_{N101D}$ and $IFN\beta_{Com}$ (Commercial nonglycosylated protein) in a cell culture growth inhibition assay. Values show mean of percent cell survival relative to untreated cells, $n>8$; Error bars show s.e.m.; $^{*}P=0.002$, $^{**}P=10^{-7}$, student's *t*-test.

modified to some extent[9]. We therefore tested whether essentially unmodified sequons could also be subject to sequence-based stabilization. For this purpose we studied the paralogous yeast proteins Ost3p and Ost6p, which are accessory subunits of OTase[7,13]. The ER lumenal domain of Ost3p has a single sequon at Asn33, which is not detectably glycosylated[35] (Fig. 4a and Supplementary Fig. 3), while Ost6p has no sequons and contains a cluster of acidic amino acids at the position corresponding to the Ost3p sequon (Supplementary Fig. 4). The ER lumenal domain of Ost6p is more easily expressed than that of Ost3p in *E. coli*[30] and shows activity in an *in vitro* peptide-binding assay when Ost3p does not[36]. To test if the higher *in vitro* activity of Ost6p relative to Ost3p correlated with increased thermodynamic stability of Ost6p, we performed guanidinium-induced equilibrium unfolding of Ost6 and Ost3 purified from *E. coli*. This showed that Ost6 ($\Delta G^0 = -79 \pm 10\,kJ\,mol^{-1}$) was more stable than Ost3 ($\Delta G^0 = -28 \pm 5\,kJ\,mol^{-1}$) (Fig. 4b,f), consistent with the presence of clusters of like-charged amino acids in place of an unused glycosylation sequon increasing thermodynamic stability and activity. To test this in the context of Ost6p, we introduced a glycosylation sequon in Ost6 at the equivalent position to the Ost3 sequon, and tested *in vitro* stability and peptide-binding activity. This variant of Ost6 with a glycosylation sequon had both reduced stability compared with wild-type Ost6 ($Ost6_{D38N}$ $\Delta\Delta G^0 = 20\,kJ\,mol^{-1}$) (Fig. 4f,g), and essentially eliminated *in vitro* peptide-binding activity (Fig. 4l,m). We next tested whether the stability and *in vitro* activity of Ost3 could be increased by targeted point mutation at the unused sequon in Ost3. We first introduced an acidic Asp in place of Asn33, which increased stability compared with wild-type protein ($Ost3_{N33D}$ $\Delta\Delta G^0 = -16\,kJ\,mol^{-1}$) (Fig. 4c), but did not result in a detectable increase in *in vitro* peptide binding (Fig. 4h,i). However, Ost3p orthologues lack sequons at this position and

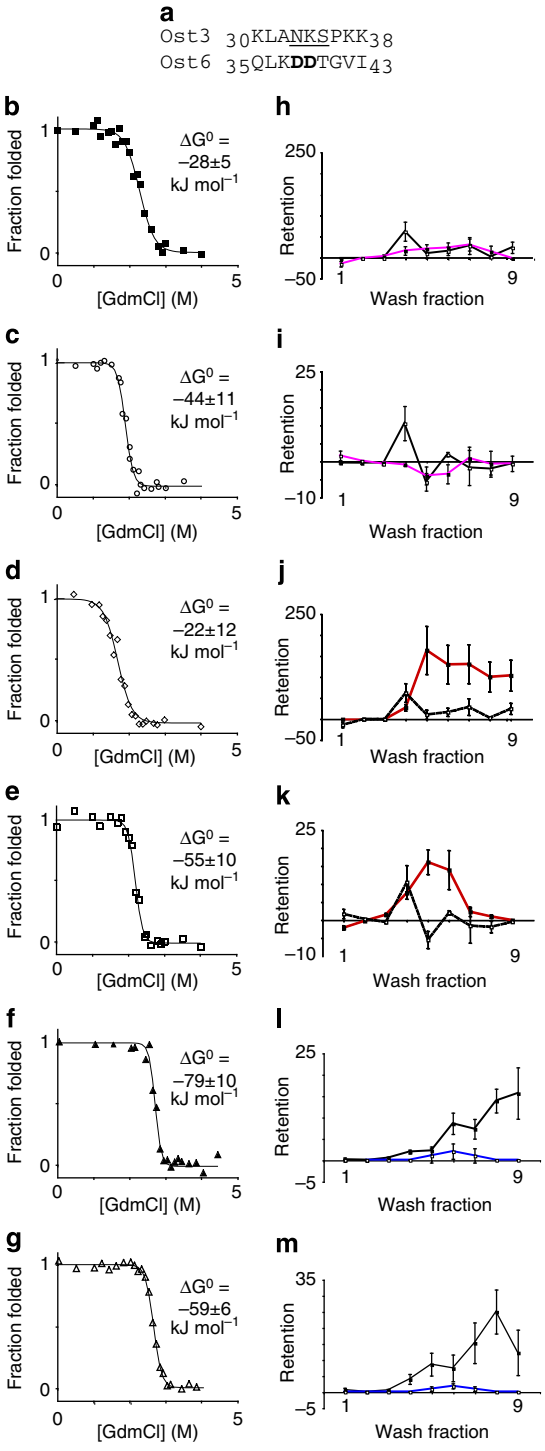


Figure 4 | Stability and function of yeast Ost3 and Ost6 variants. (**a**) Alignment of Ost3 and Ost6. The Ost3 glycosylation sequon is underlined and corresponding cluster of acidic amino acids in Ost6 are in bold. Guanidinium induced equilibrium unfolding of (**b**) Ost3, (**c**) $Ost3_{N33D}$, (**d**) $Ost3_{N33D,\ K34D}$, (**e**) $Ost3_{N33K}$, (**f**) Ost6 and (**g**) $Ost6_{D38N}$, as measured by intrinsic tryptophan fluorescence and fitted to the two-state model of protein folding. Peptide binding to Ost3 wild type (black) and $Ost3_{N33D}$ (purple) by peptides (**h**) LVIWINGDK (m/z of 1,057.6$^+$) and (**i**) FFYSNNGSQFYIR (m/z of 1,642.8$^+$), to Ost3 wild type (black) and $Ost3_{N33K}$ (red) by peptides (**j**) LVIWINGDK (m/z of 1,057.6$^+$) and (**k**) FFYSNNGSQFYIR (m/z of 1,642.8$^+$), and to Ost6 wild type (black)[36] and $Ost6_{D38N}$ (blue) by (**l**) LVIWINGDK (m/z of 1,057.6$^+$) and (**m**) AGLTFLVDLIK (m/z of 1,189.7$^+$). Values show mean of triplicates, error bars show s.e.m.

instead typically possess clusters of basic amino acids, while Ost6p orthologues have acidic residues at this position (Supplementary Fig. 4). This difference may depend on Ost3p- and Ost6-specific charge distributions on a neighbouring α-helix (Supplementary Fig. 5). We therefore tested two additional variants with clusters of either acidic or basic like-charge. While $Ost3_{N33D,K34D}$ ($_{30}$KLA**DD**SPKK$_{38}$) was less stable ($\Delta\Delta G^0 = 6\,kJ\,mol^{-1}$) (Fig. 4d), $Ost3_{N33K}$ ($_{30}$KLA**K**KSPKK$_{38}$) was more stable than $Ost3_{N33D}$ and wild-type Ost3 ($\Delta\Delta G^0 = -27\,kJ\,mol^{-1}$) (Fig. 4e). Optimal stabilization was therefore achieved through minimal mutations providing a cluster of like-charged residues. This stable $Ost3_{N33K}$ showed increased *in vitro* binding of several peptides (Fig. 4j,k). These peptides contained aromatic amino acids, complementary to Phe102 in the peptide-binding groove of Ost3, a key determinant of transient nascent polypeptide tethering by Ost3 crucial to its function in modulating OTase activity *in vivo*[30,36]. The additional thermodynamic stability of $Ost3_{N33K}$ therefore enabled its function to be investigated in this *in vitro* assay.

Discussion

We show that targeted introduction of like-charged amino acids at *N*-glycosylation sequons is a generally applicable technique for increasing the stability of normally *N*-glycosylated proteins when expressed from more efficient expression systems such as *E. coli*. The targeted nature of this approach also limits any unanticipated affects on protein structure or function of non-targeted mutagenesis. The presence of a sequon rather than a compensatory sequence at a specific location in a given protein may also depend on its protein context. For example, experimental[1,16] and computational[37] characterization of variants of CD2 adhesion domain from human (which is glycosylated) and rat (which is not glycosylated) has shown that gain of glycosylation mutations in Rat CD2 do not immediately increase protein thermodynamic stability. However, acquisition of two additional mutations proximal to the newly glycosylated Asn introduces specific glycan–protein interactions, which lead to the formation of high-occupancy hydrogen bonds and stabilize the glycosylated protein. Alternatively, while loss of glycosylation mutations in human CD2 destabilizes the protein, two additional nearby mutations rescue stability and function[38]. Experimentally, Asn-Gln site-directed mutagenesis is commonly used as a proxy for loss of glycosylation at a given site. Such a mutation may show if glycosylation at that site is important for protein folding, stability or activity, in the context of an otherwise unchanged protein sequence. However, our bioinformatics, *in vitro* and *in vivo* results here show that such a requirement could in general be easily rescued by evolutionary acquisition of a small number of stabilizing point mutations. The exact reasons why particular glycoproteins have *N*-glycosylation sites at specific locations are undoubtedly complex, depending not only on their affect on global protein stability but also on evolutionary history and the many roles of *N*-glycans in influencing protein function.

Methods

Sequence analyses. Families of yeast glycoproteins with members expressed predominantly during vegetative growth or sporulation were identified from expression data[23] and UniProt BLAST with a cutoff of e-4. Sequence analysis was performed based on a set of experimentally validated occupied glycosylation sites in *S. cerevisiae* cell-wall glycoproteins (28 sites in 13 glycoproteins)[13,28,29]. Homologues to these from *S. cerevisiae* were identified with UniProt BLAST with a cutoff of e-4, and each of these sets of homologous proteins and glycoproteins were aligned with ClustalW. For each experimentally validated glycosylation site, aligned sequences of homologous proteins were identified that did not contain a glycosylation sequon at that position (41 sequences). The characteristics of the amino-acid sequences flanking these sites (glycosylation sites, and corresponding sequences in homologues lacking glycosylation sites) were determined through a non-gapped alignment visualized with WebLogo[39] and amino-acid composition

(hydrophilic, hydrophobic, acidic or basic). pI of the amino-acid sequences was determined with ProtParam[40].

Cloning and mutagenesis. DNA encoding the ER lumenal domain of *S. cerevisiae* Gas2p as predicted in the UniProt Knowledgebase (position 25 to 531 of Gas2p/Q06135/YLR343W) was cloned into the pMAL-p2x vector, using PCR primers incorporating *Bam*HI and *Hin*dIII restriction sites, to give pMAL-GAS2. DNA encoding the GAS1 genomic locus (GAS1/P22146/YMR307W including 1 kb up and downstream) was cloned into the pRS415 vector to give pRS415-GAS1. Blunt ended DNA encoding mature MBP-Gas1 or MBP-Gas2 was amplified using PCR primers incorporating a C-terminal His-tag, phosphorylated with T4 polynucleotide kinase, and ligated into blunt ended DNA amplified from pRS415-GAS1 excluding DNA encoding mature GAS1 (ref. 41). This produced pRS415-MBP-GAS1-His and pRS415-MBP-GAS2-His, which encoded MBP-Gas1-His or MBP-Gas2-His with the native Gas1p signal sequence under the control of the native *GAS1* promoter and terminator. DNA encoding the mature protein sequence of IFNβ as predicted in the UniProt Knowledgebase (position 22 to 187 of P01574) with Cys17 changed to Ser[32] and with an N-terminal decaHis-tag, optimized for *E. coli* protein expression was synthesized and cloned into the pJexpress404 expression vector (DNA 2.0). DNA sequence encoding a C-terminal FLAG-tag was inserted in *S. cerevisiae* OST3 on plasmid YEp352 (ref. 42) using site-directed mutagenesis[41] to create pOST3-FLAG. Point mutation of GAS1, IFNβ and OST3 were performed by site-directed mutagenesis[41].

Cell culture and protein purification. BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) yeast cells and the *Δgas1* mutant derivative thereof (Open Biosystems) was transformed with pRS415, pRS415-GAS1 or variants using standard protocols. Cells were diluted to equal densities at $OD_{600\,nm}$, serially 10-fold diluted, equal volumes spotted to minimal media agar plates and incubated at 30ºC. TOP10 *E. coli* cells carrying the plasmids encoding IFNβ, Ost6 (pHis10-OST6L[43]), Ost3 (pHis10-OST3L[43]) or variants thereof were grown in LB media, and cells carrying pMAL-Gas1 (ref. 36) or pMAL-Gas2 were grown in LB-glucose media. Media was supplemented with 100 mg l^{-1} ampicillin. Protein expression was induced by addition of IPTG to 1 mM at an $OD_{600\,nm}$ of 0.7. After incubation for 4 h at 37 °C the cells were collected by centrifugation. MBP-Gas1 and MBP-Gas2 were purified as previously described[36], except that protein was eluted with 10 mM maltose. Yeast cells carrying pRS415-MBP-GAS1-His or pRS415-MBP-GAS2-His were grown in minimal media at 30 °C with shaking and collected at an $OD_{600\,nm}$ of 1.0. Cells were resuspended in 50 mM phosphate buffer pH 7.5, 150 mM NaCl, 1 × protease inhibitor cocktail (Roche) and 1 mM PMSF, and lysed by French Press. Proteins were purified using TALON resin (Clontech) and eluted in 50 mM phosphate buffer pH 7.5, 150 mM NaCl and 250 mM imidazole. The TALON resin eluate was then applied to amylose-agarose beads and purified as previously described[36], except that protein was eluted with 10 mM maltose. Disulphide-shuffled protein was produced by reduction in 6 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 with 10 mM DTT, buffer exchange into 1 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 with 0.1 mM $CuSO_4$ and incubation at 25 °C for 2 h. Protein was then buffer exchanged into 20 mM phosphate buffer pH 7.5, 200 mM NaCl and 1 mM EDTA. Collected cells expressing IFNβ variants were resuspended in 50 mM phosphate buffer pH 7.5, 150 mM NaCl, 1 × protease inhibitor cocktail (Roche) and 1 mM PMSF, and lysed by French Press. Inclusion bodies were recovered by sequential centrifugation at 18,000 rcf. The pellets were washed with 1 ml 50 mM phosphate buffer pH 7.5 and 150 mM NaCl and re-centrifuged three times. The washed pellets containing inclusion bodies were resolublized in 6 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 and 1 mM β-mercaptoethanol and incubated at 37 °C for 16 h. His-tagged IFNβ was purified using TALON resin (Clontech) and eluted in 6 M guanidinium chloride, 50 mM phosphate buffer pH 5.0 and 1 mM β-mercaptoethanol. The pH was adjusted to pH 7.5 with other constituents remaining constant, DTT was added to 20 mM and the protein was incubated at 37 °C for 1 h. Protein was buffer exchanged with PD-10 columns (GE Healthcare) into 50 mM phosphate buffer pH 7.5 and 6 M guanidinium chloride. Cysteines were oxidized to cystine by addition of 0.1 mM $CuSO_4$ and incubation at 25 °C for 2 h. Oxidation was quenched by addition of 1mM EDTA, and oxidized IFNβ was buffer exchanged with PD-10 columns into 50 mM phosphate buffer pH 7.5 and 25% glycerol at 4 °C. Cells expressing N-terminally His-tagged Ost6, Ost3 or variants thereof were resuspended in ice cold 50 mM Potassium Phosphate, pH 7.5, 150 mM NaCl, 10 mM Imidazole, 1 × complete mini EDTA free protease inhibitor (Roche) and 1 mM PMSF, and lysed by French Press. Protein was purified from clarified whole-cell extract using TALON resin and eluted with 200 mM imidazole. Proteins were buffer exchanged with PD-10 columns into 50 mM potassium phosphate buffer pH 7.0 and 1 mM EDTA. Yeast with pOST3 or pOST3-FLAG were grown in minimal media at 30 °C and collected at mid log phase. Proteins were denatured in SDS sample buffer, incubated for 1 h at 37 °C with 1 × G5 buffer in the presence or absence of Endo H[13], separated by SDS-PAGE, transferred to nitrocellulose membrane and probed with an α-FLAG antibody.

***In vitro* protein analysis.** CD spectra were obtained using a Jasco J-710 spectrometer (Jasco, Japan) as described[44]. Free energies of stabilization (ΔG^0) were determined based on the two-state model of protein folding[45] as described[30]. To determine the equilibrium transitions, intrinsic tryptophan fluorescence of proteins was measured in different concentrations of guanidinium chloride in 50 mM potassium phosphate pH 7.5 or 5.0, and 150 mM NaCl at 25 °C. ΔG^0 and cooperativity were determined from a non-linear least-squares 6 parameter fit to the data points[45], with uncertainty estimates from the corresponding six-parameter fits. Ost3 and Ost6 peptide-binding affinity chromatography was performed as described[36]. Briefly, tryptic peptides from MBP-Gas1p were applied to peptide-affinity chromatography columns containing MBP-Ost3, MBP-$Ost3_{N33K}$, MBP-Ost6 or MBP-$Ost6_{D38N}$ bound to amylose-agarose resin with or without pre-reduction with DTT. Peptides were eluted with physiological buffer, fractions collected, and peptides in the load and elution fractions were detected with MALDI-TOF-MS. Peptide retention was determined by the relative abundance of each peptide in wash fractions compared with the load, in oxidized relative to reduced. Reduction of the CxxC motif in Ost6p causes structural changes only at the peptide-binding groove[30], confirming that retained peptides bind at this groove rather than through other non-specific interactions[36].

Cell culture growth inhibition assay. A total of 4×10^3 Human MCF-7 breast tumour cells were cultured in 200 µl of phenol-free media containing 5% charcoal stripped calf serum in 96-well plates. Equal amounts of either human recombinant IFNβ (ProSpec CYT-234), IFN-β_{WT}, IFN-β_{N101D} or no protein were added to the cultures with retinoic acid (Sigma) and incubated for 4 days as described[46]. Cells were fixed with a final concentration of 3.3% (wt/vol) TCA and stained with Sulforhodamine B (Sigma R2625) as described[47], after which the OD was measured at 490 nm. A control plate to determine initial seeding intensity was included where cells were fixed 2–3 h after plating. Absorbance with this plate was set as 0% growth, and growth obtained with untreated cells was taken as 100%.

References

1. Hanson, S. R. *et al.* The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl Acad. Sci. USA* **106,** 3131–3136 (2009).
2. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–1049 (2004).
3. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126,** 855–867 (2006).
4. Rudd, P. M. *et al.* Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33,** 17–22 (1994).
5. Wormald, M. R. & Dwek, R. A. Glycoproteins: glycan presentation and protein-fold stability. *Structure* **7,** R155–R160 (1999).
6. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell. Biol.* **13,** 448–462 (2012).
7. Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16,** 47R–62R (2006).
8. Mohorko, E., Glockshuber, R. & Aebi, M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.* **34,** 869–878 (2011).
9. Schulz, B. L. in *Glycosylation*. (ed Petrescu, S.) 21–40 (Intech, 2012).
10. Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struc. Biol.* **21,** 576–582 (2011).
11. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474,** 350–355 (2011).
12. Bailey, U. M., Jamaluddin, M. F. B. & Schulz, B. L. Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J. Proteome Res.* **11,** 5376–5383 (2012).
13. Schulz, B. L. & Aebi, M. Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Mol. Cell. Proteomics* **8,** 357–364 (2009).
14. Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14,** 103–114 (2004).
15. Slynko, V. *et al.* NMR Structure Determination of a Segmentally Labeled Glycoprotein Using in Vitro Glycosylation. *J. Am. Chem. Soc.* **131,** 1274–1281 (2009).
16. Culyba, E. K. *et al.* Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science* **311,** 571–575 (2011).
17. Lizak, C., Fan, Y. Y., Weber, T. C. & Aebi, M. N-Linked Glycosylation of Antibody Fragments in Escherichia coli. *Bioconjug. Chem.* **22,** 488–496 (2011).
18. Karpusas, M. *et al.* The crystal structure of human interferon β at 2.2-Å resolution. *Proc. Natl. Acad. Sci. USA* **94,** 11813–11818 (1997).
19. Chen, M. M. *et al.* Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proc. Natl. Acad. Sci. USA* **107,** 22528–22533 (2010).
20. Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends. Biochem. Sci.* **35,** 74–82 (2010).

21. Pearse, B. R. & Hebert, D. N. Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum. *Biochem. Biophys. Acta* **1803,** 684–693 (2010).
22. Roth, J. *et al.* Protein N-glycosylation, protein folding, and protein quality control. *Mol. Cells* **30,** 497–506 (2010).
23. Chu, S. *et al.* The Transcriptional Program of Sporulation in Budding Yeast. *Science* **282,** 699–705 (1998).
24. Ragni, E., Fontaine, T., Gissi, C., Latgè, J. P. & Popolo, L. The Gas family of proteins of Saccharomyces cerevisiae: characterization and evolutionary analysis. *Yeast.* **24,** 297–308 (2007).
25. Rolli, E. *et al.* Expression, stability, and replacement of glucan-remodeling enzymes during developmental transitions in *Saccharomyces cerevisiae*. *Mol. Biol. Cell.* **22,** 1585–1598 (2011).
26. Popolo, L. *et al.* Disulfide bond structure and domain organization of yeast beta(1,3)-glucanosyltransferases involved in cell wall biogenesis. *J. Biol. Chem.* **283,** 18553–18565 (2008).
27. Hurtado-Guerrero, R. *et al.* Molecular mechanisms of yeast cell wall glucan remodeling. *J. Biol. Chem.* **284,** 8461–8469 (2009).
28. Izquiedro, L. *et al.* Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J* **28,** 2650–2661 (2009).
29. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in One: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **19,** 3758–3768 (2008).
30. Schulz, B. L. *et al.* Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA* **106,** 11061–11066 (2009).
31. Hovmoller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta. Crystallogr. D. Biol. Crystallogr.* **58,** 768–776 (2002).
32. Mark, D. F., Lu, S. D., Creasey, A. A., Yamamoto, R. & Lin, L. S. Site-specific mutagenesis of the human fibroblast interferon gene. *Proc. Natl. Acad. Sci. USA* **81,** 5662–5666 (1984).
33. Loladze, V. V. & Makhatadze, G. I. Energetics of charge-charge interactions between residues adjacent in sequence. *Proteins* **79,** 3494–3499 (2011).
34. Schweiker, K. L. & Makhatadze, G. I. Protein stabilization by the rational design of surface charge-charge interactions. *Methods Mol. Biol.* **490,** 261–283 (2009).
35. Kelleher, D. J. & Gilmore, R. The *Saccharomyces cerevisiae* oligosaccharyltransferase is a protein complex composed of Wbp1p, Swp1p, and four additional polypeptides. *J. Biol. Chem.* **269,** 12908–12917 (1994).
36. Jamaluddin, M. F. B., Bailey, U. M., Tan, N. Y. J., Stark, A. P. & Schulz, B. L. Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Sci.* **20,** 849–855 (2011).
37. Wang, X. Y., Ji, C. G. & Zhang, J. Z. Exploring the molecular mechanism of stabilization of the adhesion domains of human CD2 by N-glycosylation. *J. Phys. Chem. B* **116,** 11570–11577 (2012).
38. Wyss, D. F. *et al.* Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **269,** 1273–1278 (1995).
39. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
40. Gasteiger, E. *et al.*in *The Proteomics Handbood.* (ed Walker, J. M.) 571–607 (Humana Press, 2005).
41. Imai, Y., Matsushima, Y., Sugimura, T. & Terada, M. A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res.* **19,** 2785 (1991).
42. Spirig, U., Bodmer, D., Wacker, M., Burda, P. & Aebi, M. The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **15,** 1396–1406 (2005).
43. Mohd Yusuf, S. N., Bailey, U. M., Tan, N. Y. J., Jamaluddin, M. F. B. & Schulz, B. L. Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem. Biophys. Res. Commun.* **432,** 438–443 (2013).
44. Zalucki, Y. M., Jones, C. E., Ng, P. S., Schulz, B. L. & Jennings, M. P. Signal sequence non-optimal codons are required for the correct folding of mature maltose binding protein. *Biochim. Biophys. Acta* **1798,** 1244–1249 (2010).
45. Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **8,** 93–98 (1990).
46. Kolla, V., Lindner, D. J., Xiao, W., Borden, E. C. & Kalvakolanu, D. V. Modulation of interferon (IFN)-inducible gene expression by retinoic acid. Up-regulation of STAT1 protein in IFN-unresponsive cells. *J. Biol. Chem.* **271,** 10508–10514 (1996).
47. Vichai, V. & Kirtikara, K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat. Protoc.* **1,** 1112–1116 (2006).

Acknowledgements

This work was supported by National Health and Medical Research Council Project Grant 631615 and National Health and Medical Research Council Career Development Fellowship APP1031542 to B.L.S.

Author contributions

N.Y.T., S.C.R. and B.L.S. performed sequence analyses. N.Y.T., S.H.B and B.L.S. characterized the proteins. U.M.B. performed the growth inhibition assay. M.F.J. performed the peptide affinity chromatography assay. U.M.B. and B.L.S. wrote the manuscript with contributions from all other authors.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

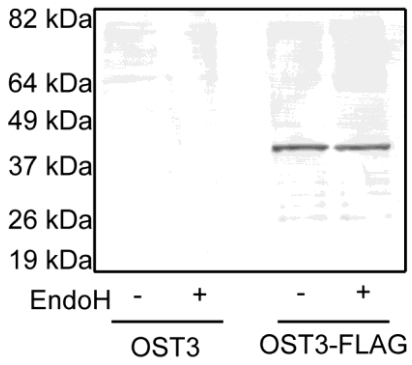
How to cite this article: Tan, N. Y. *et al.* Sequence-based protein stabilization in the absence of glycosylation. *Nat. Commun.* 5:3099 doi: 10.1038/ncomms4099 (2014).

5.4 Supplementary Information

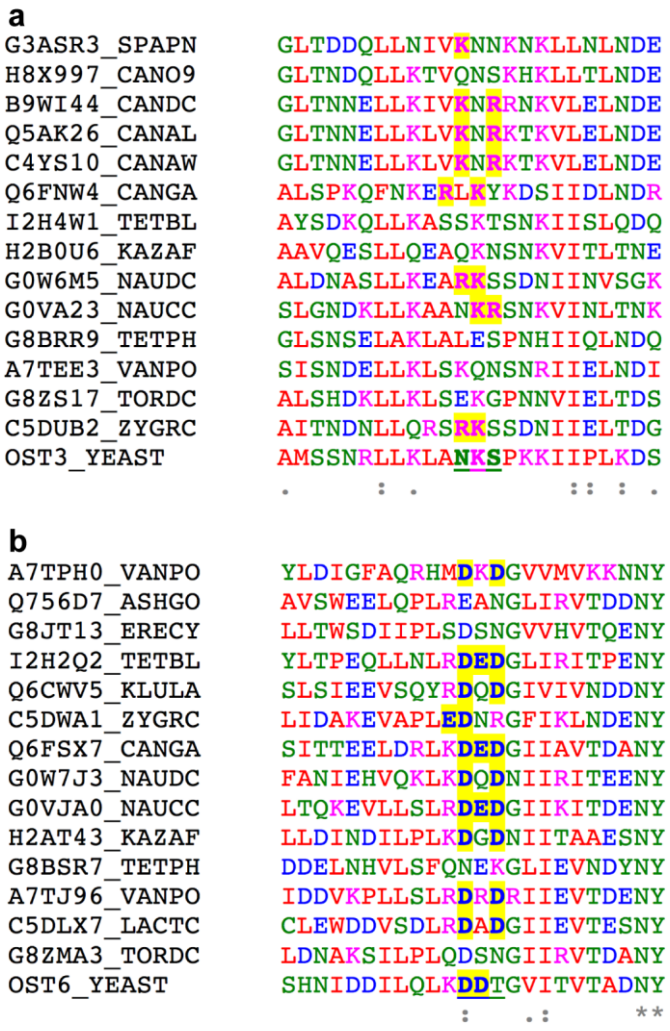
Tan NY, Bailey UM, **Jamaluddin MF**, Mahmud SH, Raman SC, Schulz BL. Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun*. **2014**, 5: 3099.

Supplementary Figure 1. Sequence alignments of *S. cerevisiae* Gas1p homologs. (**a**) Alignment of Gas1p and Gas2p surrounding the Gas1p Asn40 glycosylation site. The N-glycosylation site at Asn40 in Gas1p is underlined. Acidic amino acids at the corresponding location in Gas2p are bold and highlighted. (**b**) Multiple sequence alignment of fungal Gas1p orthologs surrounding the Gas1p Asn40 glycosylation site. The N-glycosylation site at Asn40 in Gas1p is underlined. The conserved Gly41 is bold.

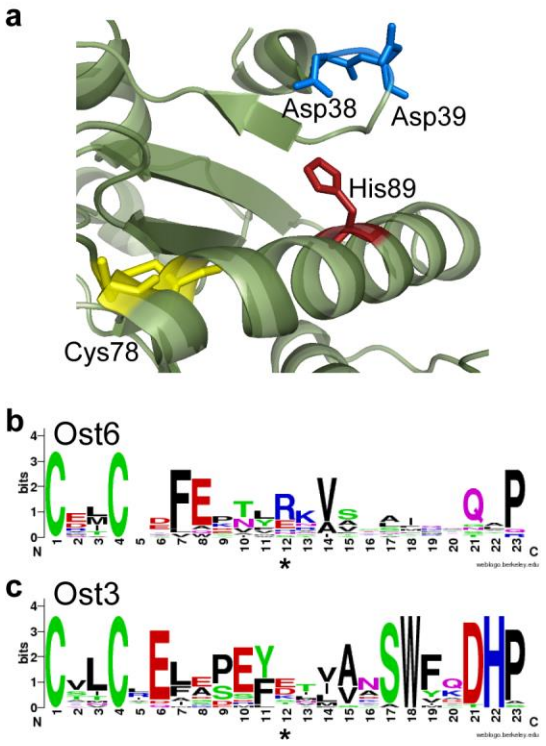
Supplementary Figure 2. Sequence alignments of human IFNβ homologs. (**a**) Multiple sequence alignment of human IFNβ paralogs. The single N-glycosylation site in IFNβ is underlined. Corresponding positions in paralogs with acidic local sequence stretches are bold and highlighted, and with a glycosylation sequon is underlined. (**b**) Multiple sequence alignment of selected mammalian IFNβ orthologs. The single N-glycosylation site in IFNβ is underlined. Corresponding position in an ortholog with a local stretch of acidic amino acids is bold and highlighted.



Supplementary Figure 3. Glycosylation of Ost3p is not detectable in *S. cerevisiae*. Whole cell extracts of *S. cerevisiae* with plasmids expressing native (OST3) or C-terminally FLAG-tagged (OST3-FLAG) Ost3p, digested with endoglycosidase H (EndoH +) or not digested (-), were separated by SDS-PAGE, blotted to nitrocellulose and probed with α-FLAG antibody.



Supplementary Figure 4. Multiple sequence alignment of selected fungal Ost3p and Ost6p orthologs. (**a**) Alignment of Ost3p orthologs. The single N-glycosylation sequon in *S. cerevisiae* Ost3p is underlined. Basic local sequence stretches at the corresponding position in orthologs are bold and highlighted. (**b**) Alignment of Ost6p orthologs. The position corresponding to the single N-glycosylation sequon in *S. cerevisiae* Ost3p is underlined in *S. cerevisiae* Ost6p. Acidic local sequence stretches at the corresponding position in orthologs are bold and highlighted.



Supplementary Figure 5. Charge complementarity between Ost3p glycosylation sequon loop and neighboring helix in Ost3p and Ost6p. (**a**) Cartoon of Ost6p crystal structure (3G7Y) [25] showing proximity of Asp38/Asp39, corresponding to the Ost3p glycosylation sequon, and His89 on a neighboring helix. (**b**) Weblogo representation of aligned sequences of this helix in fungal Ost6p and (**c**) Ost3p orthologs. The alignments begin with the Cys corresponding to Cys78 in *S. cerevisiae* Ost6p. *, Position corresponding to His89 in Ost6p.

Supplementary Table 1. Disulfide connectivity of MBP-Gas1p expressed periplasmically in *E. coli*.

m/z	*z*	M exp (Da) [^]	sum of [*]	M calc (Da) [#]	Peptides	*E. coli* C-C[a]	Yeast C-C[b]
1225.7	2	2449.3	2451.1 - 2	2449.1	SDC$_{445}$SFSGSATLQTAT TQASC$_{462}$SSALK	445-462	379-445, 372-462
1058.5	3	3172.3	1478.6 + 1695.7 - 2	3172.3	MTDYFAC$_{216}$GDDDVK, ADFYGINMYEWC$_{234}$GK	216-234	216-348, 234-265
1243.3	3	3726.9	1557.7 + 2171.0 - 2	3726.7	SYSATTSDVAC$_{348}$PATG K, NLSIPVFFSEYGC$_{265}$NEV TPR	265-348	216-348, 234-265

[^] Experimental mass of peptides observed in oxidized but not reduced tryptic digest.
* Sum of calculated peptide masses less two protons.
# calculated mass of disulfide-linked peptides.
[a] disulfide connectivity experimentally measured in *E. coli*.
[b] native disulfide connectivity previously experimentally measured in yeast.

Supplementary Table 2. Disulfide connectivity of MBP-Gas2p expressed periplasmically in *E. coli*.

m/z	*z*	M exp (Da) [^]	sum of [*]	M calc (Da) [#]	Peptides	*E. coli* C-C[a]	Yeast C-C[b]
1282.2	2	2562.4	1133.5 + 1430.6 - 2	2562.2	YEEYFSYLC$_{419}$SK, VDC$_{424}$SDILANGK	419-424	419-424
1003.6	3	3007.7	825.4 + 2184.1 - 2	3007.5	HC$_{466}$PLNDK, NVYFNLESLQPLTSESIC$_{489}$K	466-489	399-466, 392-489

[^] Experimental mass of peptides observed in oxidized but not reduced tryptic digest.
* Sum of calculated peptide masses less two protons.
# calculated mass of disulfide-linked peptides.
[a] disulfide connectivity experimentally measured in *E. coli*.
[b] native disulfide connectivity previously experimentally measured in yeast.

Chapter 6

Transient protein-protein interaction between substrate peptide and yeast Oligosaccharyltransferase subunit Ost3/6p increase *N*-glycosylation

6.1 Introduction

The ER lumenal domain of Ost3/6p proteins is predicted to have a thioredoxin-like fold and a CxxC active-site motif (Fetrow et al., 2001). Previous studies by Schulz et al. (2009) have determined the crystal structure of Ost6p that not only implicates the predicted thioredoxin-like core fold, but a conspicuous loop adjacent to the CxxC active site motif. When the CxxC motif is oxidized, the loop creates a potential peptide-binding groove, but when it is reduced, the loop becomes disordered and the groove is no longer present (Schulz et al., 2009). As sequence alignment indicates the amino acids involved in forming the peptide-binding groove are not conserved in Ost3p and Ost6p (Jamaluddin et al., 2011), it is proposed that they would recognize distinct sets of glycosylation sites on a nascent polypeptide chain (Jamaluddin et al, Unpublished work reference to Chapter 2 and 3), and sequentially increase the efficiency of glycosylation (Schulz and Aebi, 2009).

Here, we attempted to determine the crystal structure of Ost3L in complex with a physiological ligand - $Gas1p_{85\text{-}97}$ peptide (NTNVIRVYAINTT), and measure the binding affinity between them. The $Gas1p_{85\text{-}97}$ peptide was one of the physiologically relevant sequences determined in Chapter 2. It is comprised of a stretch of hydrophobic residues flanked by basic residues, and interacts directly with the peptide-binding groove of Ost3p to assist efficient glycosylation *in vivo*. This serves as a key determinant of transient nascent polypeptide tethering by Ost3p, consistent with the function of Ost3p and Ost6p in modulating *N*-glycosylation substrate selection by OTase activity *in vivo*. We also included $Ost3_{N33K}$ in this experiment, as this variant protein is a more stable form than wild-type Ost3 (Tan et al., 2014), and because this variant form of Ost3 showed increased *in vitro* peptide binding compared with wild-type Ost3. Interestingly, the peptides identified binding to $Ost3_{N33K}$ are complementary to the characteristics of the peptide binding groove (Tan et al., 2014). This transient binding of cysteine residues and hydrophobic stretches of nascent polypeptide to Ost3p/Ost6p is an optimal strategy for inhibiting local protein folding, as such hydrophobic stretches normally would be found in the internal part of folded protein domains (Schulz, 2012). This fits with the requirement of the catalytic site of OTase for unfolded or flexible protein substrate.

This chapter illustrates different approaches applied to the crystallization of the lumenal domain of Ost3p (Ost3L) with and without $Gas1p_{85\text{-}97}$ peptide. In particular, a linker strategy was utilized to increase the chance of the $Gas1p_{85\text{-}97}$ peptide binding to Ost3L. Although we were unable to obtain crystals of Ost3L-linker-$Gas1p_{85\text{-}97}$ peptide, some initial crystal hits were identified for Maltose-Binding Protein (MBP)-tagged Ost3L. Furthermore, preliminary results were obtained for the binding affinity between $Gas1p_{85\text{-}97}$ peptide and Ost3L. In general, we have piloted different

crystallization trials and conditions for setting up kinetic analysis for binding between Ost3/6L and their substrate peptides.

6.2 Materials and Methods

6.2.1 Recombinant DNA constructs

The DNA sequence corresponding to the respective ER lumenal domains of Ost3p and Ost6p (ranging from the signal peptidase cleavage sites to the beginning of the first transmembrane helix) as predicted in the UniProt Knowledgebase: residue 23-185 of Ost3p (P48439, YOR085W); residues 25-188 of Ost6p (Q03723, YML019W); and the stable form of $Ost3_{N33K}$ (Tan et al., 2014) was amplified using PCR primers and performed ligation-independent cloning (LIC) (Section 6.2.2.) into the pMCSG7 *E.coli* expression vector, producing pMCSG7-Ost3L, pMCSG7-Ost6L and pMCSG7-$Ost3_{N33K}$, respectively. This pMCSG7 vector contained an *N*-terminal His_6 tag followed by a TEV (tobacco etch virus) protease cleavage site (Stols et al., 2002). This TEV protease site is crucial for removal of His tag to improve screening and purification of Ost3/6L. For Ost3L-Flexible Linker-Peptide Sequence and $Ost3_{N33K}$-Flexible Linker-Peptide Sequence expression in *E.coli*, we used PCR primers Ost3L Fwd and Ost3-L-P Rvs to introduce peptide sequence (NTNVIRVYAINTT) from $Gas1p_{85-97}$ and flexible linker (GGSGGHMGSGG) directly *C*-Terminal of Ost3L before beginning of the first transmembrane helix, giving pMCSG7-Ost3L-Linker-peptide and pMCSG7-$Ost3_{N33K}$-Linker-peptide. We used our existing pHis10-Ost3, pHis10-Ost6 (Mohd Yusuf et al., 2013); pHis10-Ost3N33K, pHis10-Ost3-Linker-peptide and pHis10-$Ost3_{N33K}$-Linker-peptide as a template to generate all our PCR products above.

For affinity interaction assays between Ost3L/$Ost3_{N33K}$/Ost6L/Ost6Lvariants versus $Gas1p_{85-97}$, the DNA sequences of $Gas1p_{85-97}$ which contains our peptide interest sequence NTNVIRVYAINTT was amplified using PCR primers incorporating *Eco*RI and *BamH*I restriction sites and cloned into pGEX2T vector, giving pGEX2T-$Gas1p_{85-97}$. Variants with amino acid replacements (pMCSG7_$OST6L_{K96Q,K99Q}$) were constructed as described by Imai *et al* (Imai et al., 1991).

Table 6.1: Primers used for Ost3L-/ Ost6L- / $Ost3_{N33K}$ / Ost3L Linker Peptide/ $Ost3_{N33K}$ Linker Peptide

Primers	Sequences
Ost3L Fwd	5'-TACTTCCAATCCAATGCCATGTCCAGCAACAGACTACTAAAGC-3'
Ost3L Rvs	5'-TTATCCACTTCCAATGTTAGTCCATAGGTAAGTGTAAAGAGAAG-3'
Ost3-L-P-Rvs	5'- TTATCCACTTCCAATGTTAAGTGGTATTGATAGCGTAGACACGGA-3'
Ost6L Fwd	5'- TACTTCCAATCCAATGCCACTGCATCCCATAATATAG-3'
Ost6L Rvs	5'- TTATCCACTTCCAATGTTATTCCTGAACGTTGAAGGC-3'

Table 6.2: Primers used for pGEX2T-$Gas1p_{85-97}$

Primers	Sequences
pGEX2T-PepSeq-*Eco*RI Fwd	5'- ATCCGTGTCTACGCTATCAATACCACTTAAGAATTCATCGTGACTGACTG -'3
pGEX2T-PepSeq-*Bam*HI Rvs	5'- AACATTTGTGTTGGATCCACGCGGAACCAGATC-'3

6.2.2 Ligation-independent cloning

Ligation-independent cloning (LIC) was utilized for the directional cloning of PCR products without restriction enzyme digestion or ligation reactions (Aslanidis and de Jong, 1990, Haun et al., 1992). LIC-compatible vectors contain non-complimentary 12-15 base single-stranded overhangs that anneal to complementary single-stranded overhangs of the target insert. The primers amplifying the target insert have 5'- extensions so that with T4 polymerase treatment, the insert has complementary sequence to the prepared LIC vector.

The PCR products (Section 6.2.1) were analysed by agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit (Qiagen). For the generation of single-stranded overhangs, the PCR product (1 µg) was treated with T4 DNA polymerase (1 U, NEB) in the presesne of dCTP (100 mM). The pMCSG7 vectors were linearized by the blunt restriction enzyme SspI (NEB) according

to the procedure described by the manufacturer. After separating the linearized and circular vectors by agarose gel electrophoresis, the pMCSG7 vectors were extracted from the agarose gel (using a QIAquick Gel Extraction Kit (Qiagen). The linearized vector was treated with T4 DNA polymerase (NEB) as described earlier to generate the single-stranded overhangs complementary to the target site. The T4 polymerase-treated vectors (15 ng) and PCR products (25-30 ng) were mixed (incubated for 30 min) and transformed by heat shock into DH5α competent cells. Positive clones were identified with colony PCR using T7 promoter and T7 terminator primers. Positive colonies were grown in 5 ml LB media overnight and plasmids were prepared using QIAprep Spin Miniprep Kit (Qiagen).

6.2.3 Protein Expression and Purification

6.2.3.1 Large-scale expression of His-tagged-Ost3L/ His-tagged-Ost3$_{N33K}$/ His-tagged-Ost3L-Linker-peptide / His-tagged-Ost3$_{N33K}$-Linker-peptide

All proteins were expressed using the autoinduction method. Autoinduction aims to promote high density growth and protein expression. Plasmid DNA containing respectively pMCSG7-Ost3L; pMCSG7-Ost3$_{N33K}$; pMCSG7-Ost3L-Linker-Peptide; pMCSG7-Ost3$_{N33K}$-Linker-Peptide was transformed into *E.coli* DH5α expression strain by heat shock and plated onto ampicilin agar plates, which were incubated over-night at 37 ºC. Next day, a single colony was picked and inoculated into 5 ml of LB supplemented with amipicilin (100 µg/ml) and incubated overnight at 37 ºC (230 rpm) as a starter culture. For autoinduction, the overnight starter culture (2 ml) was inoculated into 535 ml autoinduction media (Studier, 2005) (ZY medium: 5g of peptone, 2.5g of yeast extract, 1M MgSO4, 50x 5052, 20x NPS) containing 100 ug/ml of ampicilin in 2 litre flask. The culture was incubated at 37 ºC (230 rpm) until the cell density reached OD_{600} around 0.6. Subsequently, the temperature was decreased to 20 ºC and the cells were grown overnight until an OD_{600} of 5-10 was reached. The cells were harvested by centrifugation at 6.500 g for 20 min at 4ºC and the cell pellets were either processed immediately or stored at -80 ºC. Due to time constraints, I selected pMCSG7-Ost3$_{N33K}$ and pMCSG7-Ost3$_{N33K}$-Linker-peptide construct as my main priority for polyhistidine purification and affinity assay or crystallisation set up.

6.2.3.2 Purification of His_6-tagged of $Ost3_{N33K}$ and $Ost3_{N33K}$-Linker-peptide

The cell pellets were resuspended in the lysis buffer containing ice cold 20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM PMSF, 30 mM immidazole, 1x Complete Mini EDTA free protease inhibitor (Roche Diagnostics) and 100 µg/ml DNase at the ratio of 5 ml lysis buffer per gram of cell pellets. Cells were lysed using a sonicator (Sonifier W-450D; Branson) and the lysates were subject to centrifugation at 15,000 g at 4 ºC for 30 mins. The supernatant was collected and injected onto a pre-washed HisTrap^{TM} FF 5 ml column (GE Healthcare) with lysis buffer (20 mM of Tris-HCl pH 8, 500 mM NaCl, 30 mM immidazole) at 2.5 ml/min. The column was washed with 10 column volumes of Washing Buffer (20 mM of Tris-HCl pH8, 500 mM NaCl, 40 mM immidazole) at 4 ml/min to remove unbound bacterial proteins. Bound proteins containing either His-tagged-$Ost3_{N33K}$ or His-tagged-$Ost3_{N33K}$-Linker peptide were eluted with 5 column volumes of Elution Buffer (20 mM of Tris-HCl pH8, 500 mM NaCl, 250 mM immidazole). All purification steps were performed at 4 ºC. The identity and final purity of $Ost3_{N33K}$ and $Ost3_{N33K}$-Linker peptide were confirmed by SDS-PAGE (data not shown). All the fractions containing the protein of interest were pooled and concentrated to 2 ml followed by subsequent dilution to 4 -column volumes of TEV cleavage buffer (50 mM Tris-HCl pH 8, 250 mM NaCl and 0.5 mM EDTA) to dilute the immidazole concentration. The His_6-tag was removed by overnight treatment with TEV protease (100 µl, 8 mg/ml) at 4 ºC in the TEV cleavage buffer. After cleavage, the protein was injected onto a pre-equilibrated His Trap^{TM} FF 5 ml column (GE Healthcare) to remove uncleaved protein, His_6-tag, the TEV protease (His-tagged) and other contaminating proteins. The flow through was collected and concentrated to 5 ml using Amicon-Ultra centrifugal filter devices (10 kDa cutoff) and injected on to a HiLoad 26 / 60 Superdex 200 PG column (GE Healthcare) pre-equilibrated with the gel filtration buffer (20 mM Tris HCl pH 8, 125 mM NaCl). The peak fractions were pooled, concentrated (10-15 mg/ml), snap-frozen as 50 µl aliquots in liquid nitrogen and stored at-80ºC.

6.2.3.3 Purification of GST-tagged $Gas1_{85\text{-}97}$ peptide

The cell pellets were resuspended in the lysis buffer (5 ml of lysis buffer for 1 gram of cell pellets) containing 50 mM Tris HCl pH 8, 125 mM NaCl, 1 mM EDTA and 2 mM DTT. Resuspended cells were lysed by sonification (Sonifier W-450D; Branson) at 4 ºC and cell debris was removed by centrifugation at 15,000 g for 30 minutes at 4 ºC (Avanti J-26; Beckman Coulter). The soluble supernatant fractions were collected and purified by IMAC (immobilized metal affinity chromatography) using a 5 ml GST (Glutathione S-transferase) column (GST^{TM} FF 5 ml, GE Healthcare), where the soluble fractions were loaded onto the GST column for protein binding at

2.5 ml/min. To remove non-specifically bound bacterial proteins, the GST column was washed using the lysis buffer until the absorbance reached a steady state at 4 ml/min. GST-tagged protein was eluted by elution buffer containing 10 mM glutathione in the lysis buffer. The identity and final purity of GST-tagged $Gas1_{85-97}$ were confirmed by SDS-PAGE.

6.2.3.4 Purification of MBP-Ost3L

The cell pellets of MBP-Ost3Lwere resuspended in lysis buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl and 1 mM EDTA, 1 mM PMSF and 100 µg/ml DNase) at the ratio of 5 ml lysis buffer per 1 g of cell pellets. Then, the cell pellets were lysed at 4 ºC by sonication (Sonifier W-450D; Branson). The lysates were centrifugated at 15,000 g for 30 min at 4 ºC (Avanti J-26; Beckman Coulter) and the supernatant was collected and applied onto the MBP Trap HP 5 ml column (GE Healthcare) at 2.5 ml/min. The column was further washed with column buffer (50 mM Tris HCl pH 7.5, 150 mM NaCl and 1 mM EDTA) to about 50-100 ml at 4 ml/min to remove non-specifically bound proteins. Finally, the MBP-Ost3L protein was eluted with 40 ml of elution buffer containing column buffer and 10 mM maltose (D-(+) maltose monohydrate). The fractions containing MBP-Ost3L verified by SDS-PAGE were pooled and reduced volume to 5 ml for injecting to HiLoad 26 /60 Superdex 75 PG column (GE Healthcare). The elution based on the peak fractions from the FPLC were pooled together and concentrated (by using 10 kDa cutoff Amicon-Ultra centrifugal filter device) to 5 ml of gel filtration. The concentrated protein (77 mg from 5 ml) was snap-frozen as 25 µl aliquots in liquid nitrogen and stored at -80 ºC and ready for crystallisation tray setup.

6.2.4 Crystallization screening and optimization

For initial crystallization screening, several commercially available sparse matrix screens were set up using the Mosquito robot (TTP LabTech, UK) at the UQ ROCX Diffraction Facility. The screens used in this study were: Index (Hampton Research), PEGIon (Hampton Research), JCSG+ (Molecular Dimensions) and Proplex (Molecular Dimensions). The droplets were set up with equal ratio of protein and reservoir solution (100 nl of protein and 100 nl of reservoir solution) and were equilibrated against 100 µl of reservoir solution. The plates were stored in an automated imaging system (Rock Imager) at 20 ºC and each drop was imaged after 0, 1, 2, 3, 5, 8, 13 and 21 days. The conditions of the hits from the initial crystallization screens were optimized in a 24-well plate format by varying the precipitant concentrations and the pH values of buffer. Each droplet containing 1 µl of protein and 1 µl of reservoir solution was equilibrated against 500 µl of reservoir solution at 18 ºC incubator.

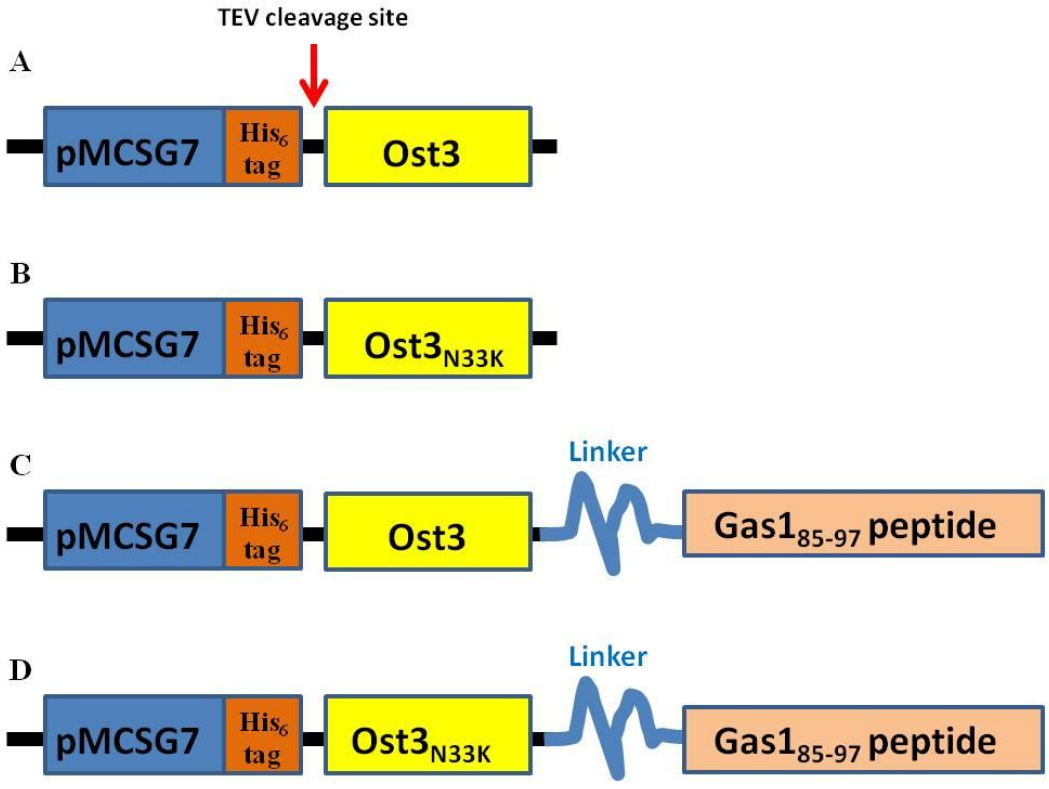
6.2.5 Kinetic analysis of ligand-analyte interaction

We performed kinetic analysis of ligand-analyte interaction between GST-Gas1$_{85-97}$ peptide and purified Ost3$_{N33K}$ (Section 5.2.3.2). We firstly used anti-GST antibody biosensor to capture the ligand: GST-Gas1$_{85-97}$ peptide and then measure its binding to different concentrations of Ost3$_{N33K}$ from 300 µM, 200 µM, 150 µM, 100 µM and 0 µM. We also performed reverse analysis using Ni-NTA biosensor captured His$_6$-tagged Ost3$_{N33K}$ first, and then measured its binding to GST-Gas1$_{85-97}$ peptide from 672 µM, 67.2 µM and 0 µM. There are five steps in this kinetic analysis: 1) Initial Baseline, 2) Loading, 3) Baseline, 4) Association and 5) Dissociation as instructed by manufacturer (ForteBio BLItz System). It is essential to hydrate the biosensors for at least 10 min with experimental buffer (gel filtration buffer: 20 mM Tris HCl pH 8, 125 mM NaCl) before use. At the first step, the biosensor was incubated with 250 µl of experimental buffer to obtain a stable baseline for 30 sec. At the Loading step, the biosensor was dipped into the drop holder filled with 4 µl of the ligand for 120 secs. Then, the third step is to wash away unbound ligand with 250 µl of experimental buffer for 30 sec. During the Association step, the biosensor loaded with ligand was dipped into 4 µl of analyte in the drop holder to measure interaction between ligand and analyte. Lastly at the Dissociation step, the biosensor was incubated with 250 µl of experimental buffer for 120 sec. Data was collected in real-time fashion and analysed using 1:1 binding mode. Negative controls were also performed to ensure there was no binding between GST and Ost3$_{N33K}$.

6.3 Results

6.3.1 Rationale of construct design and cloning of pMCSG7-Ost3L/ pMCSG7-Ost3$_{N33K}$/ pMCSG7-Ost3L-Linker-Peptide / pMCSG7-Ost3$_{N33K}$-Linker-Peptide

To maximize the chances of successful crystallization using this technique, initial construct design is crucial. The length of the flexible linker region consisted of eleven amino acids (GGSGGHMGSGG) with minimal hydrophobicity to avoid unnecessary aggregation between the protein of interest Ost3L and the peptide sequence (NTNVIRVYAINTT) from Gas1p$_{85-97}$ peptide. Based on the published structure of Ost6L (Schulz et al., 2009), we used PyMOL to estimate that the distance between the *C*-terminus of Ost3L directly before the beginning of the first transmembrane helix and the Ost3L peptide-binding groove was approximately 44.2Å. We predicted that a peptide linker of this length should be sufficient to allow Gas1p$_{85-97}$ peptide to access and bind at the targeted Ost3L peptide-binding groove. Due to my limited PhD time, I selected pMCSG7-Ost3$_{N33K}$ and pMCSG7-Ost3$_{N33K}$-Linker-peptide construct as my main priority for polyhistidine purification and affinity assay or crystallisation set up. Ost3$_{N33K}$ was preferred than Ost3 wild type because Ost3$_{N33K}$ binds strongly to the peptide as shown from Tan *et al* (2014).



Figures 6.1: Scheme expression of the pMCSG7 vector harbouring Os3 or Ost3$_{N33K}$ with and without linker peptide. Using the designated primers described above (6.2.1), all PCR product of

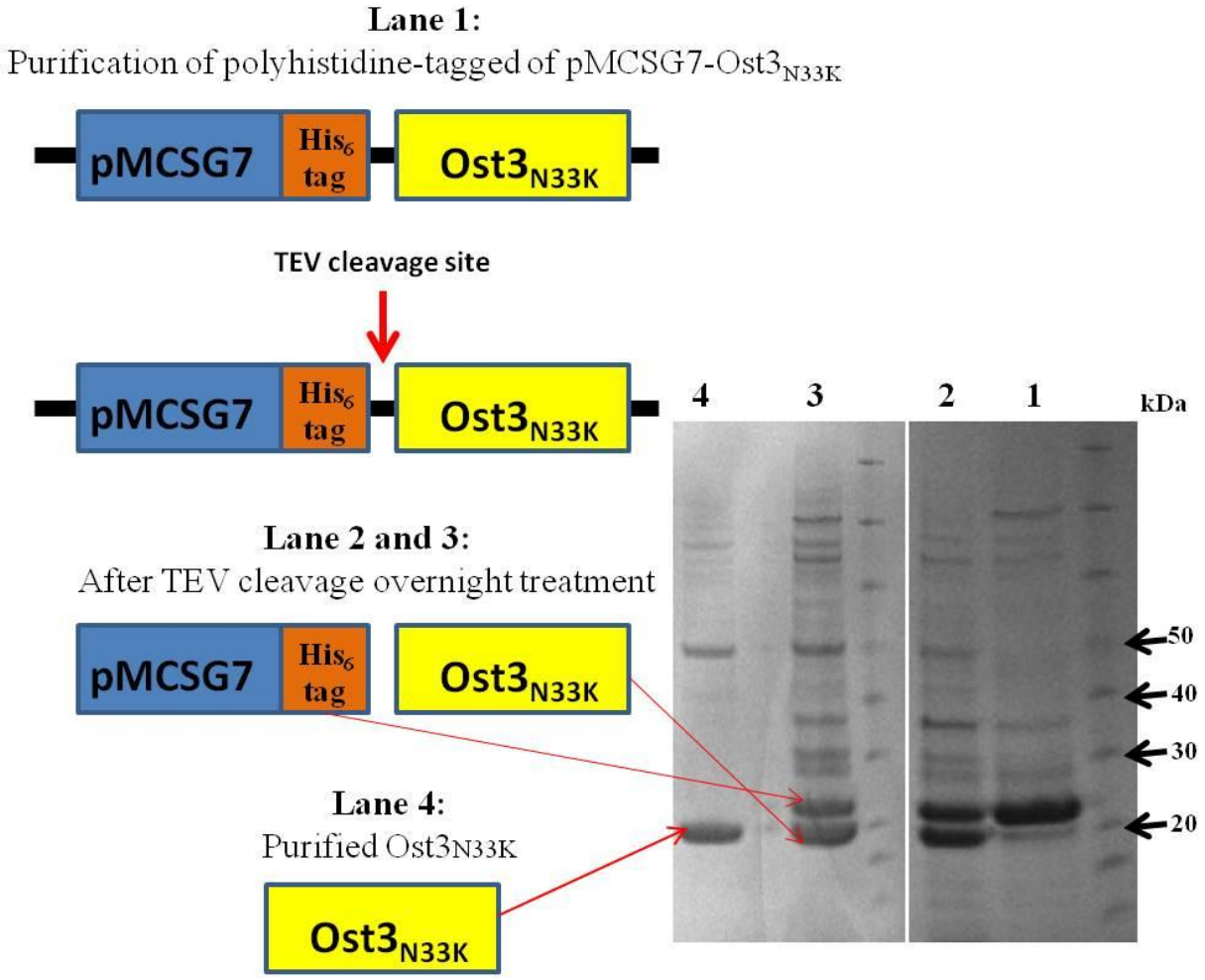
A) pMCSG7-Ost3L, B) pMCSG7-Ost3$_{N33K}$, C) pMCSG7-Ost3L-Linker-Gas1$_{85-97}$ peptide and D) pMCSG7-Ost3$_{N33K}$-Linker-Gas1$_{85-97}$ peptide had the correct size insert through ligation-independent cloning (LIC) (Section 6.2.2.) into the pMCSG7 *E.coli* expression vector. This pMCSG7 vector contained an *N*-terminal hexahistidine (His) tag followed by a TEV protease cleavage site (Stols et al., 2002). This TEV protease site is crucial for removal of His$_6$ tag to improve crystallization screening. To make Ost3L-Linker-Gas1$_{85-97}$ peptide and Ost3$_{N33K}$- Linker-Gas1$_{85-97}$ peptide expression construct, we used PCR primers Ost3L Fwd and Ost3-L-P Rvs to introduce flexible linker and Gas1$_{85-97}$ peptide sequences at the C terminal end of Ost3L.

6.3.2 Purification of Ost3$_{N33K}$ and Ost3$_{N33K}$-Linker-Gas1$_{85-97}$ peptide

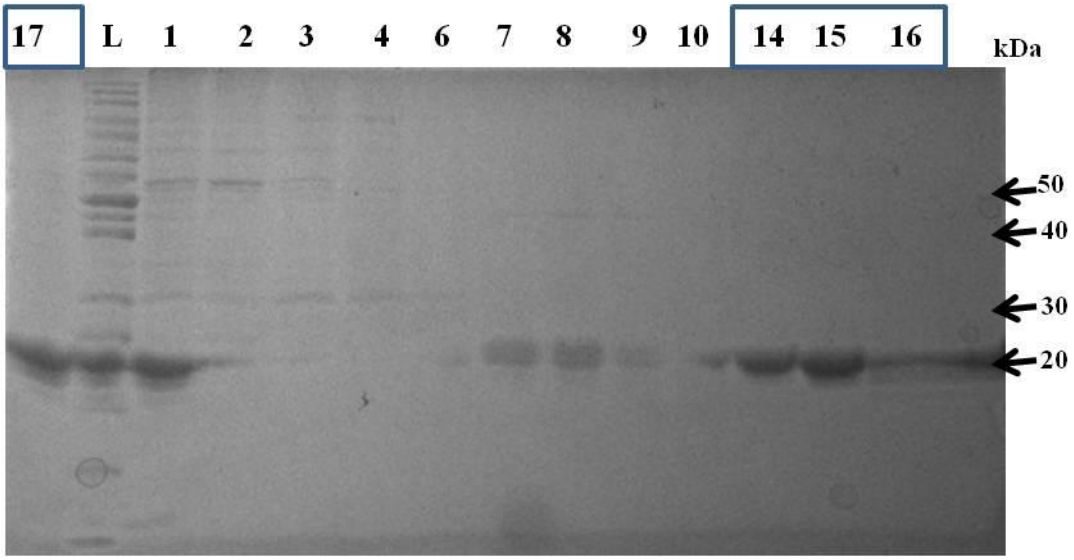
Here the aim was to achieve the expression and purification of Ost3$_{N33K}$ and Ost3$_{N33K}$-linker-Gas1$_{85-97}$ peptide in *E.coli* BL21(DE3) expression, and performed polyhistidine tag purification, and finally TEV cleavage digestion to remove His tag for obtaining the purified protein. For purification of Ost3$_{N33K}$ (Figure 6.2a), the distinct bands in lane 1 showed the confirmed purified proteins of Ost3$_{N33K}$ using a polyhistidine tagged purification. Lane 2 and 3 indicate clear separation of a band from His$_6$ tag after overnight treatment with TEV protease. After cleavage, the protein was injected onto a pre-equilibrated His Trap^{TM} FF 5ml column to remove un-cleaved protein, the His$_6$ tag, the TEV protease and other contaminating proteins. A small aliquot from the elution was collected and ran on the protein gel to check for purified Ost3$_{N33K}$ (Lane 4). The purified protein of Ost3$_{N33K}$ was collected and concentrated for crystallization set up and affinity assays.

For purification of pMCSG7-Ost3$_{N33K}$-Linker-Gas1$_{85-97}$ peptide (Figure 6.2b), after TEV cleavage, the protein was injected into the FPLC to determine protein purity. We collected the elution fractions from 1 to 17 from the FPLC and identified the distinct bands fractions 14 - 17 (Label in Blue). We pooled and concentrated fractions for crystallization purpose.

We trialed this purified Ost3$_{N33K}$ (Figure 6.2A) and Ost3$_{N33K}$-Linker-Gas1$_{85-97}$ peptide (Figure 6.2B) which was earlier cleaved from His$_6$ tag in different conditions through several commercially available sparse matrix screens but none of them produced any crystals.



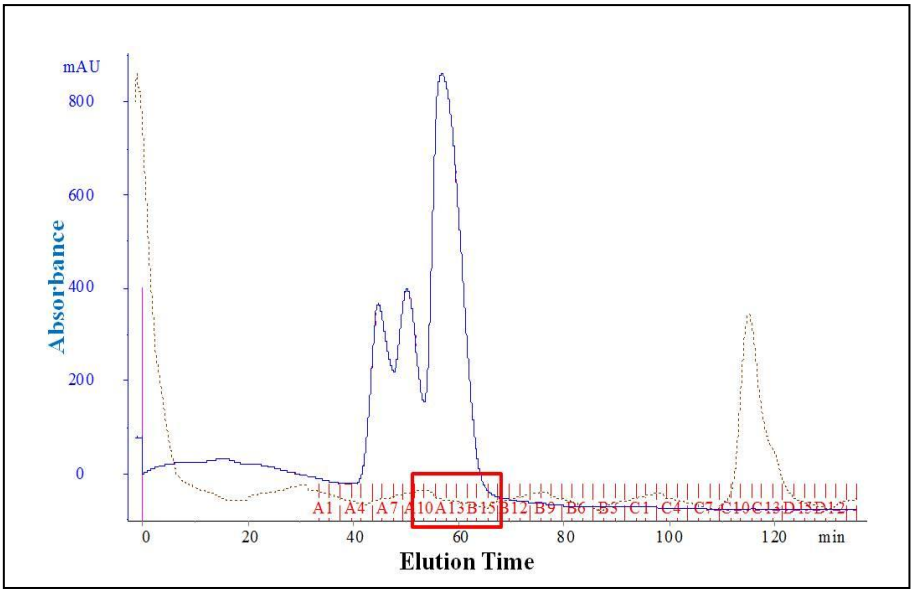
Figures 6.2a: Purification of $Ost3_{N33K}$; Lane 1: purification of His_6-tagged of $Ost3_{N33K}$, Lane 2 and 3: After TEV overnight digestion separating $Ost3_{N33K}$ from His_6 Tag and Lane 4: Purified $Ost3_{N33K}$. On the right, it is the cartoon representation of the purification steps for $Ost3_{N33K}$. The red arrows indicated the respective inserts.



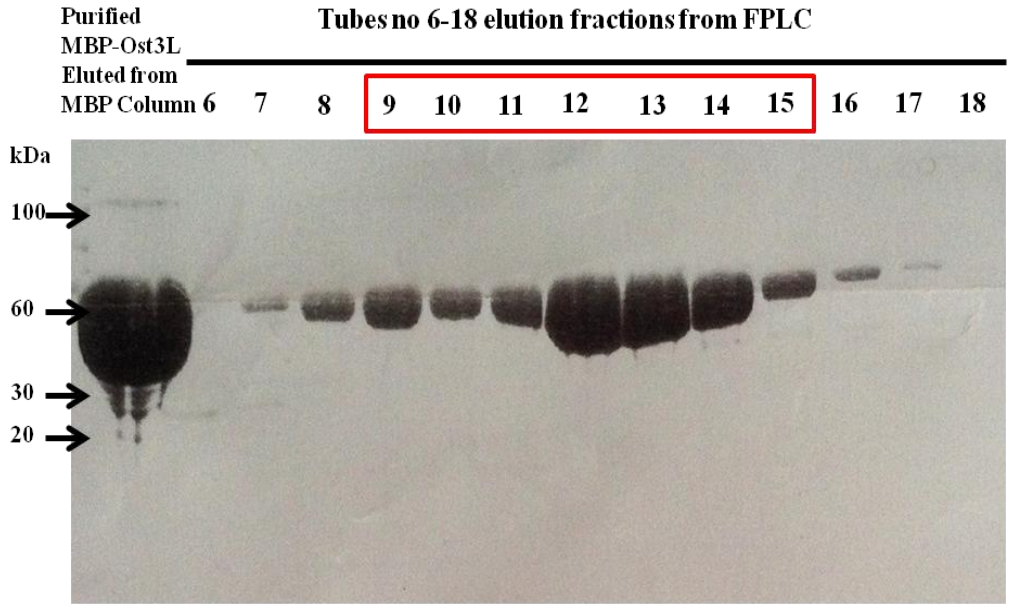
Figures 6.2b: Purification of Ost3$_{N33K}$-Linker-peptide; The lanes 14 - 17 (in blue box) were corresponding to the peak fractions from the FPLC. These fractions were pooled together and concentrated for crystallization.

6.3.3 Purification of MBP-Ost3L

We used our existing MBP-Ost3L (Jamaluddin et al., 2011) to facilitate crystallization. The cells were grown in large scale of 3 litre cell culture using the autoinduction method. The protein was purified using affinity to MBP amylose resin followed by injection into FPLC and gel filtration (Section 6.2.3.4). Suitable fractions of the elution based on the peak fractions from the FPLC (Figure 6.3A) were pooled together and concentrated (by using 10 kDa cutoff Amicon-Ultra centrifugal filter device). The concentrated protein (77 mg from 5 ml) was snap-frozen in 25 µl aliquots in liquid nitrogen and stored at -80 ºC ready for crystallisation tray setup.



(A)

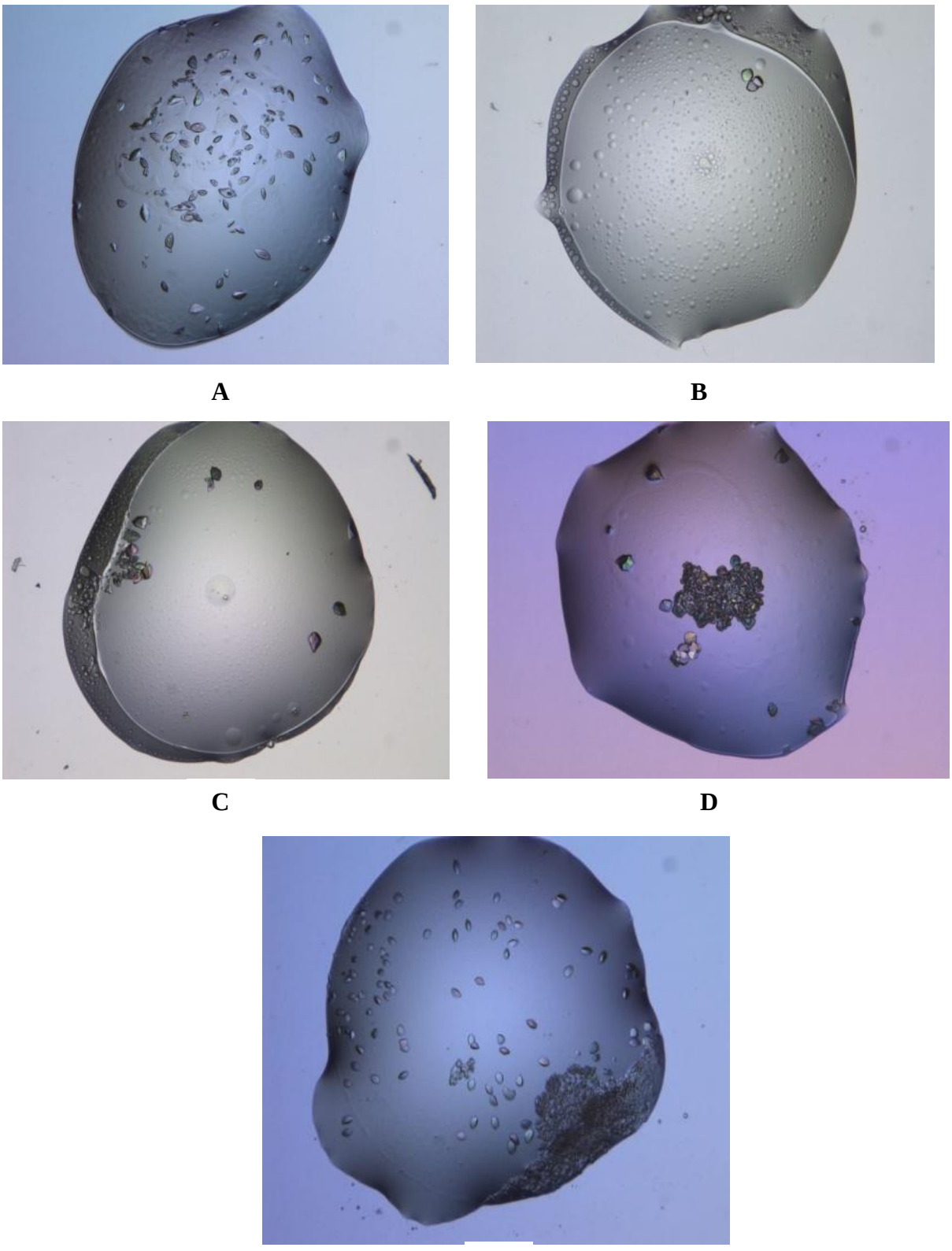


(B)

Figure 6.3: Gel filtration and Purification of MBP-Ost3L (A) Gel filtration peak fractions profile of MBP-Ost3L; fractions 9 - 15 (in red box) of the elution peak fractions from the FPLC were pooled together and concentrated. **(B)** Purification of the MBP-Ost3L construct agreed well with the theoretical weights of MBP-Ost3L (61.2 kDa).

6.3.4. Crystallization screening of MBP-Ost3L

For setting up crystallization screens, a suitable protein concentration (20 mg/ml) was determined first using Qubit Protein concentration measurements. Here we utilized two commercially available sparse matrix screens which is Index (Hampton Research) and PEGIon (Hampton Research) to identify conditions that could induce crystallization. All the screens were set up using the hanging drop method with 100 nl protein and 100 nl reservoir solution per drop (Section 6.2.4). A promising hit was indentified for MBP-Ost3L protein; crystals formed after 5-8 days of setting up the drop and was observed in 4 out of 99 wells in Index plate, reference to 1) Index well no 45 with following condition: 25% w/v Polyethylene glycol 3,350 and 0.1 M Tris HCl pH 8.5; 2) Index well no 62: 0.2 M Trimethylamine N-oxide dihydrate, 20% w/v polyethylene glycol monomethyl ether and 0.1 M Tris HCl pH 8.5; 3) Index well no 73: 0.2 M NaCl, 25% w/v Polyethylene glycol 3,350 and 0.1 M Tris HCl pH 8.5 and 4) Index well no 90: 20% w/v Polyethylene glycol 3,350 and 0.2 M Sodium Formate. For PEGIon plate, 1 hit was indentified for MBP-Ost3L out of 99 wells reference to this condition in 0.1 M Tris HCl pH 8 and 28% w/v Polyethylene glycol 4,000 in PEGIon well C9. In order to optimize the condition, customized grid screens were designed, in which both the pH and precipitant concentration were changed in the higher as well as lower direction from initial condition. However, so far I have not been able to reproduce the original crystals. $Ost3_{N33K}$, $Ost3_{N33K}$-Linker-peptide, pHis10-Ost3-Linker-peptide and pHis10-$Ost3_{N33K}$-Linker-peptide were initially trialed for crystallization in different conditions through several commercially available sparse matrix screens but none of them produced any crystals.



A **B**

C **D**

E

Figure 6.4: Images of crystals of MBP-Ost3L after 8 days in the A) Index well no 45, B) Index well no 62. C) Index well no 73, D) Index well no 90, E) PEGIon well C9.

6.3.5. Purification of GST-Gas1$_{85-97}$ and GST pull-down of Ost3$_{N33K}$ protein

For affinity assay measuring the strength of interaction between Ost3$_{N33K}$ and a short stretch of peptide sequence ($_{85}$NTNVIRVYAINTT$_{97}$) in Gas1p, we have successfully cloned Gas1$_{85-97}$ in the pGEX2T vector and expressed this fusion protein in *E.coli* strain BL21 (DE3) by the autoinduction method in large scale (1litre cell culture). We chose this peptide in relevance to our results in the Chapter 2 which showed that this Gas1$_{85-97}$ peptide requires the neutral peptide binding groove of Ost3p for efficient glycosylation. The GST-Gas1$_{85-97}$ protein was successfully purified (Figure 6.5: Lane 3) using a 5 ml GST column (GST^{TM} FF 5 ml, GE Healthcare) followed by gel filtration (Section 6.2.3.3). The GST tag improved the solubility of the peptide so we were able to obtain 6 mg of purified protein from 1 litre culture. The purity of proteins was analysed by SDS-PAGE analysis below. We also performed GST pull down to capture our purified Ost3$_{N33K}$ proteins (Lane 2) that interact with GST-Gas1$_{85-97}$peptide (Lane 3). Our results showed a weak faint band (red box) in lane 6 which suggested there is a binding between GST-Gas1$_{85-97}$ with Ost3$_{N33K}$ peptide binding groove.

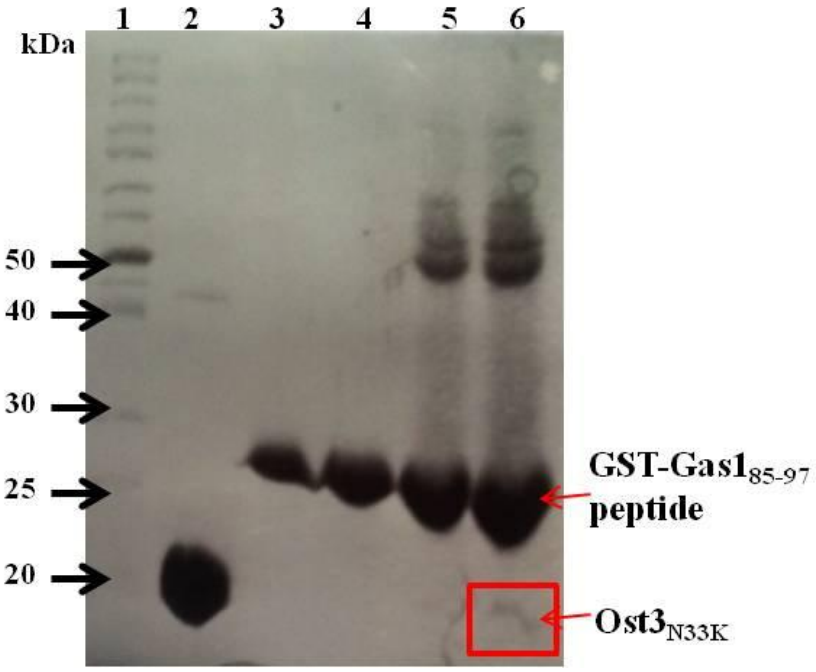


Figure 6.5: Purification of Ost3$_{N33K}$ and GST-Gas1$_{85-97}$: Lane 1) molecular weight ladder, Lane 2) Ost3$_{N33K}$, Lane 3) GST-Gas1$_{85-97}$ peptide, Lane 4) Gas1$_{85-97}$ peptide, Lane 5) GST-Gas1$_{85-97}$ peptide versus Ost3$_{N33K}$ with 4 times dilution and Lane 6) GST-Gas1$_{85-97}$peptide versus Ost3$_{N33K}$ with 1 time dilution.

6.3.6. Kinetic analysis of $Ost3_{N33K}$ versus GST-$Gas1_{85-97}$ peptide interaction

In this experiment we performed kinetic analysis of ligand-analyte interactions where the GST-$Gas1_{85-97}$ peptide which acts as the ligand was first loaded in biosensors on the BLItz system followed by monitoring the association and dissociation of $Ost3_{N33K}$ to the ligand (Figure 6.6). We use the high affinity anti-GST antibody on biosensor that binds to our GST-$Gas1_{85-97}$ peptide. The aim here is to determine that $Ost3_{N33K}$ can bind to GST-$Gas1_{85-97}$ peptide captured on biosensor. We tested using the purified and concentrated $Ost3_{N33K}$ (6 mg/mL) which was cleaved using the TEV protease to remove the His_6 tag and prepared different concentration of $Ost3_{N33K}$. In the first run (Green colour), we loaded the GST-$Gas1_{85-97}$ peptide on the biosensor and no $Ost3_{N33K}$ being replaced by an experimental buffer. Our results showed a clear indication of green signal which suggest no binding during association phase observed between 180 to 300 seconds. In the third run (Orange colour), we loaded the GST-$Gas1_{85-97}$ on the biosensor and subsequently loaded 300 µM of $Ost3_{N33K}$. As expected, we saw the spectra at the highest peak during association phase with a K_D affinity of interaction of 0.5 mM. In the subsequent fourth run (Purple colour), fifth run (Grey colour) and sixth run (Black colour), we loaded different concentrations of 100 µM, 150 µM and 200 µM of $Ost3_{N33K}$ to their respective runs versus GST-$Gas1_{85-97}$ peptides (150 nM) on the biosensor. From the spectra, we saw a very clear difference and binding during association phase when we compared the first run to the increasing concentration of $Ost3_{N33K}$, indicating specific concentration dependent binding between $Ost3_{N33K}$ and GST-$Gas1_{85-97}$ peptide.

We also performed reverse analysis using the anti Ni-NTA antibody on biosensor to bind to our His_6-tagged $Ost3_{N33K}$ (Figure 6.7). We performed three runs of experiments with different concentrations of our GST-$Gas1_{85-97}$ peptide, and one run (labelled in orange) with GST-$Gas1_{85-97}$ peptide replaced with the experimental buffer and applied these samples to our Anti-His antibody biosensor bound $Ost3_{N33K}$. For Run 2, we performed ten times dilution of the GST-$Gas1_{85-97}$ peptide from Run 1. From the spectra, we saw a very clear difference during association phase when we compared the first run and the last run, indicating specific binding between $Ost3_{N33K}$ and the peptide with a K_D affinity of interaction of approximately 2 Molar.

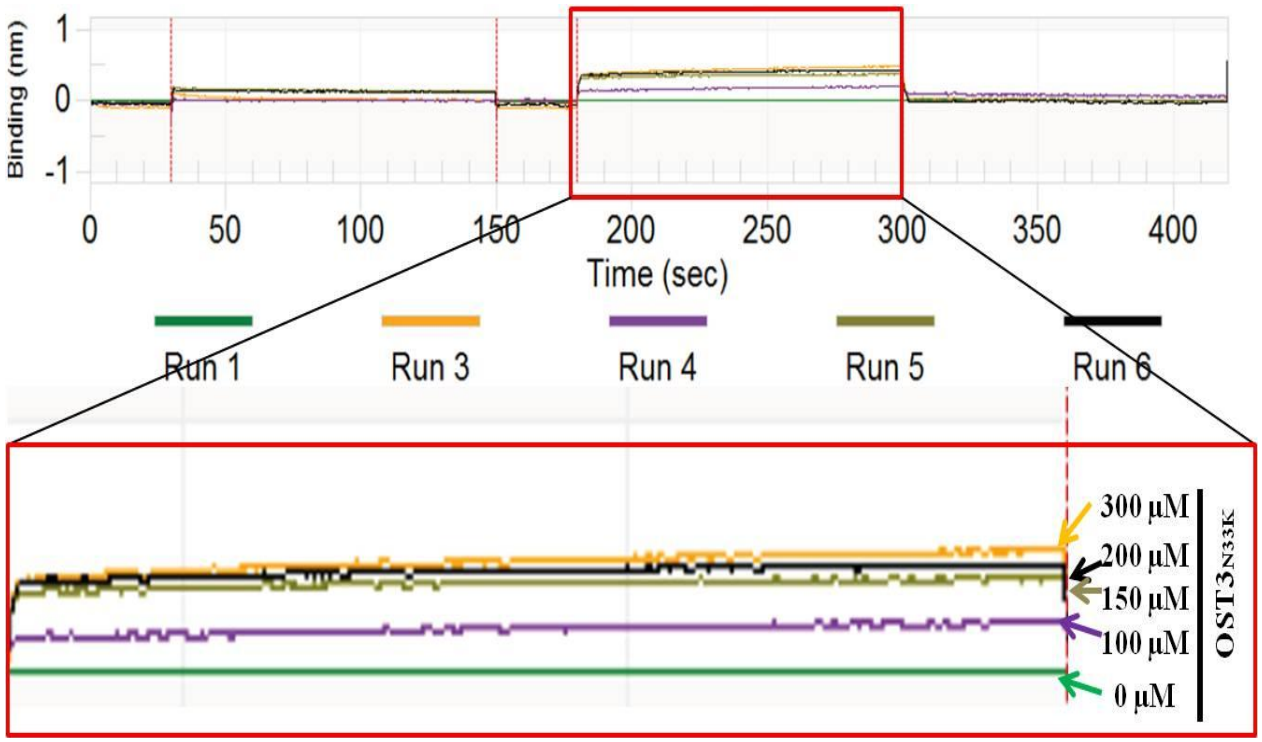


Figure 6.6: Kinetic analysis of GST-Gas1$_{85-97}$ peptide (ligand) versus different concentrations consisting of 300 µM (Orange), 200 µM (Black), 150 µM (Grey), 100 µM (Purple) and 0 µM (Green) of Ost3$_{N33K}$ (analyte) interactions. Sensorgram displayed the data collected in real time. Y axis: binding signal in nanometer (nm) unit; X-axis: analysis time in sec unit.

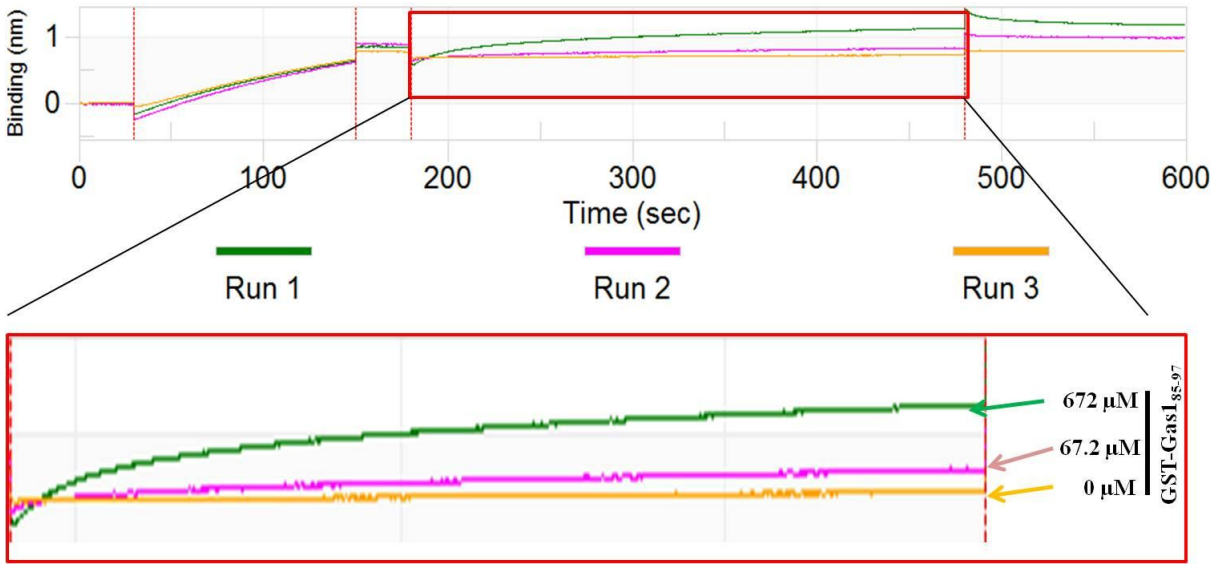


Figure 6.7: Kinetic analysis of Ost3$_{N33K}$ (ligand) versus different concentrations consisting of 672 µM (Green), 67.2 µM (Pink) and 0 µM (Orange) of GST-Gas1$_{85-97}$ peptide (analyte) interactions. Sensorgram displayed the run data collected in real time – binding signal was measured in nanometers (nm) as a function of time (sec).

6.4 Summary

The main objective in the first part of this chapter was to make protein expression constructs of $Ost3_{N33K}$ and $Ost3_{N33K}$-linker-$Gas1_{85-97}$ peptide in pMCSG7 vector using LIC cloning method, and further obtain high yield of purified proteins for crystallographic purposes. As experiments using constructs $pHis_{10}$-Ost3-Linker-$Gas1_{85-97}$ peptide and $pHis_{10}$-$Ost3_{N33K}$-Linker-$Gas1_{85-97}$ peptide failed to produce any crystals in any commercial crystallization screens (not shown), we questioned if either the 10 histidine residues tag or flexible linker or peptide caused aggregation or interfered with the formation of protein crystal contacts. Consequently, we used the pMCSG7 vector that provides an *N*-terminal hexahistidine (His) purification tag followed by a TEV protease cleavage site (Stols et al., 2002). The TEV protease cleavage site is crucial for the removal of His_6 tag to reduce the flexibility of proteins and improve chances of obtaining crystals. Our own experiments have demonstrated that the expression of pMCSG7 $Ost3_{N33K}$ was successfully obtained. This was also established by SDS PAGE results (Figure 6.2A) which confirmed the purification of the cut and uncut version of $Ost3_{N33K}$ from the His_6tag. The purified $Ost3_{N33K}$ was collected and concentrated for crystallization and affinity assays. The commercial crystallization screens were set up for both purified $Ost3_{N33K}$ and $Ost3_{N33K}$-Linker-$Gas1_{85-97}$ peptide with His_6 tag removed. However, none of these produced crystals. Additionally, these proteins became unstable and aggregated during the process of concentration which was not improved by adding 5% glycerol in the purification buffer. This caused a major hurdle for obtaining milligram quantities of highly purified, soluble protein for crystallization screens. Due to this issue and limited time for my PhD, we used our existed MBP Ost3L construct using pMAL vector (Jamaluddin et al., 2011) to produce MBP tagged Ost3L for crystallization. The usage of the MBP tag is to improve the solubility of protein and chances on crystallization. By doing so we obtained five promising conditions for crystallizing MBP-Ost3L protein from two commercial screens: Index and PEGIon (Hampton Research). Although we have not been able to reproduce the crystals in the initial grid screens, some techniques for promoting crystallization will be utilized in future work such as stripe- or macro-seeding.

The second part of this chapter illustrates kinetic analysis of the binding between $Ost3_{N33K}$ and GST-$Gas1_{85-97}$ peptide. The preliminary result indicated a weak binding between $Ost3_{N33K}$ and $Gas1_{85-97}$ peptide which is consistent with the proposed mechanistic model of Ost3/6p function in which the binding has to be transient to allow translocation and folding of the glycosylated polypeptide to precede (Figure 1.6). Further repetition of the kinetic experiments is required to ensure statistical integrity. Furthermore, oxidized and reduced controls are very essential to validate and confirm that peptides bind to the peptide binding groove, which only present in the oxidized

Ost3, rather than non-specifically elsewhere on Ost3. This concept is relevant to previous studies that have highlighted the ability of the active-site loop of Ost6L to generate a peptide-binding groove during its oxidized state but not in the reduced state (Schulz et al., 2009). We have also tried to have a similar peptide sequence synthesized with a fluorophore coupled and measured the binding of the peptide to Ost3p using fluorescence anisotropy (not shown). Unfortunately, the reaction mixture aggregated, possibly due to the high hydrophobicity of the peptide. This is the main reason we chose to proceed with a fusion protein for further studies. Alternatively, we would use well-established ELISA assays to determine the binding affinity of Ost3/6L and peptide. For this we would need to clone Ost3/Ost6p and variants into the pET-30a vector (Novagen), which tags the proteins with an S-tag which can be recognised with horseradish peroxidise (HRP) (Kaneshiro et al., 2002). By supplying the HRP substrate that can be catalysed by HRP, the binding can be measured by detecting the absorbance change at A420nm (Matsuura et al., 2003, Lange et al., 2010, Takeda et al., 2011, Marfori et al., 2012).

In conclusions, the structure of Ost3 bound to a peptide would allow the precise interactions to be known, which could then be tested *in vitro* and *in vivo*. The binding affinity assays could be used to test these interactions *in vitro*. Finally, the quantitative comparison of *in vitro* peptide binding by Ost3p and Ost6p would give insights into how these subunits function *in vitro*, especially concerning the balance between their specificity and ability to complement each other.

6.5 References

ASLANIDIS, C. & DE JONG, P. J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res,* 18**,** 6069-6074.

FETROW, J. S., SIEW, N., DI GENNARO, J. A., MARTINEZ-YAMOUT, M., DYSON, H. J. & SKOLNICK, J. 2001. Genomic-scale comparison of sequence- and structure-based methods of function prediction: Does structure provide additional insight? *Protein Science,* 10**,** 1005-1014.

HAUN, R. S., SERVENTI, I. M. & MOSS, J. 1992. Rapid, reliable ligation-independent cloning of PCR products using modified plasmid vectors. *Biotechniques,* 13**,** 515-8.

IMAI, Y., MATSUSHIMA, Y., SUGIMURA, T. & TERADA, M. 1991. A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res,* 19**,** 2785.

JAMALUDDIN, M. F. B., BAILEY, U.-M., TAN, N. Y. J., STARK, A. P. & SCHULZ, B. L. 2011. Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Science,* 20**,** 849-855.

KANESHIRO, E. S., ROSENFELD, J. A., BASSELIN-EIWEIDA, M., STRINGER, J. R., KEELY, S. P., SMULIAN, A. G. & GINER, J.-L. 2002. The Pneumocystis carinii drug target S-adenosyl-L-methionine:sterol C-24 methyl transferase has a unique substrate preference. *Mol Microbiol,* 44**,** 989-999.

LANGE, A., MCLANE, L. M., MILLS, R. E., DEVINE, S. E. & CORBETT, A. H. 2010. Expanding the definition of the classical bipartite nuclear localization signal. *Traffic,* 11**,** 311-23.

MARFORI, M., LONHIENNE, T. G., FORWOOD, J. K. & KOBE, B. 2012. Structural basis of high-affinity nuclear localization signal interactions with importin-alpha. *Traffic,* 13**,** 532-48.

MATSUURA, Y., LANGE, A., HARREMAN, M. T., CORBETT, A. H. & STEWART, M. 2003. Structural basis for Nup2p function in cargo release and karyopherin recycling in nuclear import. *EMBO J,* 22**,** 5358-69.

MOHD YUSUF, S. N., BAILEY, U. M., TAN, N. Y., JAMALUDDIN, M. F. & SCHULZ, B. L. 2013. Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem Biophys Res Commun,* 432**,** 438-43.

SCHULZ, B. L. 2012. Beyond the sequon: sites of *N*-glycosylation. *inGlycosylation ed Petrescu S. 21-40Intech*.

SCHULZ, B. L. & AEBI, M. 2009. Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific *N*-Glycosylation Efficiency. *Molecular & Cellular Proteomics,* 8**,** 357-364.

SCHULZ, B. L., STIRNIMANN, C. U., GRIMSHAW, J. P. A., BROZZO, M. S., FRITSCH, F., MOHORKO, E., CAPITANI, G., GLOCKSHUBER, R., GRÜTTER, M. G. & AEBI, M. 2009. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p

defines site-specific glycosylation efficiency. *Proceedings of the National Academy of Sciences*.

STOLS, L., GU, M., DIECKMAN, L., RAFFEN, R., COLLART, F. R. & DONNELLY, M. I. 2002. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. *Protein Expr Purif,* 25**,** 8-15.

STUDIER, F. W. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif,* 41**,** 207-34.

TAKEDA, A. A., DE BARROS, A. C., CHANG, C. W., KOBE, B. & FONTES, M. R. 2011. Structural basis of importin-alpha-mediated nuclear transport for Ku70 and Ku80. *J Mol Biol,* 412**,** 226-34.

TAN, N. Y., BAILEY, U.-M., JAMALUDDIN, M. F., MAHMUD, S. H. B., RAMAN, S. C. & SCHULZ, B. L. 2014. Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun,* 5.

Chapter 7

Conclusion and Future Directions

7.1 Conclusion

Asparagine (*N*)-linked glycosylation is an abundant post-translational modification of many proteins in eukaryotes that is important in diverse biological systems (Schwarz and Aebi, 2011). The *N*-linked glycans perform a multitude of functions at a cellular and organismal level (Mohorko et al., 2011). This process is catalyzed by the multiprotein complex oligosaccharyltransferase (OTase) (Kelleher and Gilmore, 2006, Mohorko et al., 2011, Schwarz and Aebi, 2011, Schulz, 2012). Alterations in the biosynthesis of the unique lipid-linked oligosaccharide substrate of OTase therefore affect the glycosylation efficiency and functionality of many different proteins and results in diverse human diseases (Mohorko et al., 2011). While catalysis is performed by the Stt3p protein subunit of OTase (Nilsson et al., 2003, Yan and Lennarz, 2005, Lizak et al., 2011), our knowledge about the functions of the additional protein subunits of OTase in the *N*-glycosylation process are not clearly defined.

In Chapter 2, we established that the homologs Ost3p/Ost6p subunits of yeast OTase directly interact with polypeptide substrate to enhance glycosylation of asparagines proximal and *C*-terminal to sequestered sequences. Previous studies have shown that incorporation of either Ost3p or Ost6p into OTase results in two enzyme isoforms with different protein substrate preferences *in vivo*, and that *in vitro* Ost3p/Ost6p transiently bind peptides with complementary characteristics to their peptide-binding grooves (Schulz and Aebi, 2009, Schulz et al., 2009, Jamaluddin et al., 2011). In this experiment, we performed yeast genetics and glycoproteomics to show that this peptide-binding activity of Ost3p/Ost6p is required for site-specific glycosylation by OTase *in vivo*. This confirmed for the first time that there are transient direct physical interactions between nascent polypeptide and Ost3p/Ost6p. However, in some cases the presence of basic residues in both nascent polypeptide and Ost3p/Ost6p caused electrostatic repulsion, limiting binding and inhibiting glycosylation at particular sites. This finding presented us with an experimental tool to probe the details of the interactions between nascent polypeptide and Ost3p/Ost6p that impact site-specific glycosylation. We performed scanning basic-to-neutral mutagenesis through a model substrate protein, and monitored the effect of each mutation on sitespecific glycosylation occupancy. This showed that binding of substrate polypeptide to Ost3p/Ost6p enhanced glycosylation at sites proximal and *C*-terminal to the sequestered sequence. We then validated selected physiologically relevant peptide substrates *in vitro*.

Based on our results from this integrated study using *in vivo* yeast genetics, glycoproteomics and *in vitro* biochemistry, we presented a model of the mechanisms by which Ost3p/Ost6p increase the efficiency of site-specific *N*-glycosylation. In this model, transient sequestration of stretches of

nascent polypeptide by Ost3p/Ost6p at the translocon separates newly translocating polypeptide from pre-translocated polypeptide. Glycosylation sequons in the short flexible loops of nascent polypeptide formed by this mechanism are therefore presented to the active site of OTase in a glycosylation competent form. This is a surprising and novel insight into the mechanisms of cotranslocational polypeptide modification in the endoplasmic reticulum.

In Chapter 3, we developed new methodologies using Selected Reaction Monitoring Mass Spectrometry coupled to peptide affinity chromatography which provided significant improvements in sensitivity and quantification that defined the peptide binding characteristics of native yeast Ost3p and Ost6p. We showed that Ost3p and Ost6p bind distinct subsets of peptides from a yeast cell wall glycoprotein substrate, with Ost3p binding peptides rich in aromatic acids and Ost6p binding peptides rich in aliphatic amino acids. Our data supports a model of Ost3p and Ost6p function in which they transiently bind complementary hydrophobic stretches of nascent polypeptide substrate to optimally inhibit protein folding, thereby increasing glycosylation efficiency at nearby asparagine residues. In the following Chapter 4, we showed an improved analytical method for site-specific glycosylation analysis using PNGase F to label glycosylation sites with an asparagines-aspartate conversion that creates a new endoproteinase AspN cleavage site, followed by proteolytic digestion, and detection of peptides and glycopeptides by LC-ESI-MS/MS (Bailey et al., 2012). This approach will be useful in site-specific glycosylation analysis in many model systems and clinical application. Overall, these two novel analytical methods will provide advanced knowledge for improved detection and quantification of glycosylation sites *in vivo* and peptide interactions *in vitro*.

In Chapter 5, we described a novel strategy for engineering glycoprotein stability which we used to determine the detailed function of Ost3p and Ost6p using *in vitro* peptide affinity chromatography. In previous studies, we have shown that the ER lumenal domain of Ost6 is more easily expressed than Ost3p in *E.coli* and shows *in vitro* peptide binding activity when Ost3 does not (Jamaluddin et al., 2011). Interestingly, Ost3p has a single glycosylation sequon, while Ost6p has no glycosylation sequons and contains a cluster of acidic amino acids at the position corresponding to the Ost3p sequon. While *N*-glycosylation is vital for the folding, stability and function of many glycoproteins, it is common that in the homologs of such glycoproteins the precise sites of *N*-glycosylation are not conserved. This suggests that a properties of the sequence in a protein can at least rescue for lack of *N*-glycosylation. Indeed, we successfully showed that incorporation of targeted like-charged amino acids at *N*-glycosylation sequons in Ost3/6p increases *in vitro* stability and activity (Tan et al., 2014). When we performed optimal stabilization through a minimal number of local point mutations at the Ost3p glycosylation sequon, the resulting protein showed increased *in vitro* binding

of peptides, which were complementary to the characteristics of the peptide binding groove of Ost3, a key determinant of transient nascent polypeptide tethering by Ost3p. This is consistent with the function of Ost3p and Ost6p in modulating *N*-glycosylation substrate selection by OTase activity *in vivo*. This compensation would allow flexibility in the evolution of the precise sites of *N*-glycosylation in glycoproteins, and could also be used to engineer normally glycosylated proteins so that they can be expressed as non-glycoproteins in efficient bacterial expression systems.

In Chapter 6, we extended our investigation using the best expressing and most stable variants of Ost3L from Chapter 5 and performed crystallization trials with a view to obtain a 3D crystal structure of Ost3L bound with a physiological substrate peptide. Different strategies were applied to the crystallization of Ost3L with and without peptide. Again, the peptide sequences would be physiologically relevant sequences as determined from Chapter 2. Although we were unable to obtain crystals of Ost3L-linker-Gas1p$_{85\text{-}97}$ peptide, some initial crystal hits were identified for MBP-tagged Ost3L. We also performed initial experiments to measure the binding affinity between Ost3L and peptide and detected weak but specific binding. These preliminary results are in agreement with the proposed mechanistic model of Ost3/6p function in *N*-glycosylation (Figure 1.6) in which this binding has to be transient to allow translocation and folding of the glycosylated polypeptide to precede (Schulz et al., 2009).

In summary, this thesis has made a significant contribution towards our understanding of the molecular mechanisms of glycosylation site selection by OTase, the connection between glycoprotein folding and *N*-glycosylation in the endoplasmic reticulum, and opens up potential applications in synthetic glycobiology such as engineering of glycosylation sites and OTases. Our data provides advance information towards solving the model of the mechanisms by which Ost3/6p increase the efficiency of site-specific *N*-glycosylation. Here we detailed the mechanism of this regulatory system through mapping the precise sites of interactions between Ost3/6p and Gas1p nascent polypeptide which was not previously identified. This data allowed us to present a model, in which we propose that short hydrophobic stretches of substrate polypeptide transiently bind to Ost3p or Ost6p, leading to the formation of a short flexible loop of new translocated polypeptide. Glycosylation sequons in the short flexible loops of nascent polypeptide formed by this mechanism are therefore presented to the active site of OTase in a glycosylation competent form. This is a novel insight into the mechanisms of co-translocational polypeptide modification in the endoplasmic reticulum. Our study also give insights into how the evolution of OTase isoforms has increased the substrate range of efficiently glycosylated proteins; how flexibility in the presence and position of *N*-glycosylation sites can allow sequence evolution; and how this processes can be used for creating engineered cellular systems for glycoprotein expression and engineering glycoproteins

so that they can be expressed even in bacteria that do not glycosylate. These findings will be of general interest and utility to a wide variety of researchers across the fields of glycobiology, protein biochemistry, structural biology, protein expression, biopharmaceutical development and protein evolution.

7.2 Future Directions

A key feature of the lumenal domain of the Ost3/6 subunits is their peptide-binding groove, which in Ost6p is oxidation state dependent, and their oxidoreductase activity (Schulz et al., 2009). In the present study, we have addressed primarily the peptide-binding properties of these subunits using reverse genetics. However, it would be of interest to know if the mutations introduced into the peptide-binding sites also affect the redox potential of the oxidoreductase. If this is the case, the oxidation state dependent formation of the peptide binding domain would be affected and this would result in altered peptide binding as well. It is evident that such an explanation does not explain the effect of mutations in the substrate but it is an alternative explanation for the severe phenotype of the Ost3 mutations.

Another aspect of this study worthy of future investigation would be to investigate and characterize the human homologs of Ost3p and Ost6p - MagT1 and TUSC3. While roles of MagT1 and TUSC3 regulating *N*-glycosylation or magnesium transport have been proposed, the mechanisms of MagT1/TUSC3 involvement is these processes are not clear. We already have the yeast double deletion strain (Δ*ost3*/Δ*ost6*) from Chapter 2, which are the yeast homologs of MagT1/TUSC3 encoding OTase accessory subunits. We could also create a strain with genomic deletion of *ALR1*, encoding the major yeast magnesium transporter. Previous studies have shown that Δ*alr1* yeast requires magnesium supplementation for growth and Δ*ost3*/Δ*ost6* shows inefficient *N*-glycosylation at many sites in diverse glycoproteins (Zhou and Clapham, 2009). We could then test if expression of human MagT1 or TUSC3 can rescue either or both of these phenotypes. This would provide clear evidence of a direct role for MagT1 and/or TUSC3 in magnesium transport, regulation of *N*-glycosylation, or both.

Alternatively, whole-genome sequencing of patient DNA may reveal mutations in other subunits of OTase, and the subsequent characterization of the *N*-glycoproteome will provide additional information to define the molecular function of the different OTase components (Mohorko et al., 2011). Eventually, this understanding of OTase-mediated catalysis will provide insights for the development of drugs that affect the properties of the enzyme. In addition, glycoproteins are also

significant as valuable therapeutic products (Lawrence and Betenbaugh, 2002). About 70% of the new drug candidates that have been brought to commercialisation are glycosylated proteins (Durocher and Butler, 2009). A number of industrially important glycoproteins are used as drugs, including erythropoietin, interferon, colony stimulating factors, and blood-clotting factors, and this has led to an upsurge in glycosylation engineering. By genetically engineering glycosylation pathway in different hosts, this aims to optimise glycoproteins to achieve desirable biological, biochemical or physical properties. A detailed understanding of the molecular mechansisms by which OTase controls site-specific *N*-glycosylation will facilitate efficient targeted and systems glycoengineering.

7.3 References

BAILEY, U.-M., JAMALUDDIN, M. F. & SCHULZ, B. L. 2012. Analysis of Congenital Disorder of Glycosylation-Id in a Yeast Model System Shows Diverse Site-Specific Under-glycosylation of Glycoproteins. *Journal of Proteome Research,* 11**,** 5376-5383.

DUROCHER, Y. & BUTLER, M. 2009. Expression systems for therapeutic glycoprotein production. *Curr Opin Biotechnol,* 20**,** 700-7.

JAMALUDDIN, M. F. B., BAILEY, U.-M., TAN, N. Y. J., STARK, A. P. & SCHULZ, B. L. 2011. Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Science,* 20**,** 849-855.

KELLEHER, D. J. & GILMORE, R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology,* 16**,** 47R-62R.

LAWRENCE, S. M. & BETENBAUGH, M. J. 2002. Addressing Insect Cell Glycosylation Deficiencies Though Metabolic Engineering. *In:* AL-RUBEAI, M. (ed.) *Cell Engineering.* Springer Netherlands.

LIZAK, C., GERBER, S., NUMAO, S., AEBI, M. & LOCHER, K. P. 2011. X-ray structure of a bacterial oligosaccharyltransferase. *Nature,* 474**,** 350-5.

MOHORKO, E., GLOCKSHUBER, R. & AEBI, M. 2011. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *Journal of Inherited Metabolic Disease,* 34**,** 869-878.

NILSSON, I., KELLEHER, D. J., MIAO, Y., SHAO, Y., KREIBICH, G., GILMORE, R., VON HEIJNE, G. & JOHNSON, A. E. 2003. Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J Cell Biol,* 161**,** 715-725.

SCHULZ, B. L. 2012. Beyond the sequon: sites of N-glycosylation. *inGlycosylation ed Petrescu S. 21-40Intech.*

SCHULZ, B. L. & AEBI, M. 2009. Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Molecular & Cellular Proteomics,* 8**,** 357-364.

SCHULZ, B. L., STIRNIMANN, C. U., GRIMSHAW, J. P. A., BROZZO, M. S., FRITSCH, F., MOHORKO, E., CAPITANI, G., GLOCKSHUBER, R., GRÜTTER, M. G. & AEBI, M. 2009. Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proceedings of the National Academy of Sciences.*

SCHWARZ, F. & AEBI, M. 2011. Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struct Biol,* 21**,** 576-82.

TAN, N. Y., BAILEY, U.-M., JAMALUDDIN, M. F., MAHMUD, S. H. B., RAMAN, S. C. & SCHULZ, B. L. 2014. Sequence-based protein stabilization in the absence of glycosylation. *Nat Commun,* 5.

YAN, A. & LENNARZ, W. J. 2005. Unraveling the mechanism of protein N-glycosylation. *J Biol Chem,* 280**,** 3121-4.

ZHOU, H. & CLAPHAM, D. E. 2009. Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development. *Proc Natl Acad Sci U S A,* 106**,** 15750-5.

7.4 Appendix

Other author's publication work which contributed substantially to a manuscript entitled "Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p" published in *Biochemical and Biophysical Research Communications*. The full details of publication as follow:

Mohd Yusuf SN, Bailey UM, Tan NY, **Jamaluddin MF**, Schulz BL. (2013). Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem Biophys Res Commun*. 432(3):438-443.

Biochemical and Biophysical Research Communications 432 (2013) 438–443

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Mixed disulfide formation *in vitro* between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p

Siti N.H. Mohd Yusuf, Ulla-Maja Bailey, Nikki Y. Tan, Muhammad Fairuz Jamaluddin, Benjamin L. Schulz *

School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Brisbane, Queensland 4072, Australia

ARTICLE INFO

Article history:
Received 22 January 2013
Available online 12 February 2013

Keywords:
N-glycosylation
Saccharomyces cerevisiae
Thioredoxin
Disulfide
Mass spectrometry

ABSTRACT

Oligosaccharyltransferase (OTase) glycosylates selected asparagine residues in secreted and membrane proteins in eukaryotes, and asparagine (N)-glycosylation affects the folding, stability and function of diverse glycoproteins. The range of acceptor protein substrates that are efficiently glycosylated depends on the action of several accessory subunits of OTase, including in yeast the homologous proteins Ost3p and Ost6p. A model of Ost3p and Ost6p function has been proposed in which their thioredoxin-like active site cysteines form transient mixed disulfide bonds with cysteines in substrate proteins to enhance the glycosylation of nearby asparagine residues. We tested aspects of this model with a series of *in vitro* assays. We developed a whole protein mixed disulfide interaction assay that showed that Ost6p could form mixed disulfide bonds with selected cysteines in pre-reduced yeast Gas1p, a model glycoprotein substrate of Ost3p and Ost6p. A complementary peptide affinity chromatography assay for mixed disulfide bond formation showed that Ost3p could also form mixed disulfide bonds with cysteines in selected reduced tryptic peptides from Gas1p. Together, these assays showed that the thioredoxin-like active sites of Ost3p and Ost6p could form transient mixed disulfide bonds with cysteines in a model substrate glycoprotein, consistent with the function of Ost3p and Ost6p in modulating N-glycosylation substrate selection by OTase *in vivo*.

1. Introduction

Asparagine (N-)linked glycosylation is a common co- and post-translational modification that affects the structure and function of secretory and membrane bound proteins in eukaryotes [1], archaea [2] and some bacteria [3]. In eukaryotes, nascent polypeptide is N-glycosylated by the enzyme oligosaccharyltransferase (OTase) after entry into the endoplasmic reticulum (ER) lumen through the translocon [4–6]. N-glycans enhance the efficiency of glycoprotein folding in the ER by increasing the solubility of folding nascent polypeptide and by recruiting the disulfide isomerase ERp57 through the lectins calnexin and calreticulin [7–10]. Glycan structures are modified during glycoprotein traffic through the Golgi, and the final structures present on mature glycoproteins can play important roles in regulating protein function [1,11]. Changes in the structure of glycans attached to proteins are also associated with numerous diseases [12–15].

OTase catalyses the key step of N-glycosylation, transfer of oligosaccharide from a dolicholpyrophosphate carrier to asparagines in nascent polypeptides translocated into the ER lumen [16]. OTase is a multiprotein complex consisting of a catalytic subunit, Stt3p, and varying numbers of accessory protein subunits in different organisms [5,17,18]. The acceptor peptide-binding site of Stt3p specifically recognizes asparagine residues in glycosylation sequons (Asn-Xaa-Ser/Thr; Xaa≠Pro) [19], substantially increasing the efficiency of glycosylation at these sites [20]. The accessory subunits are also important for determining the protein acceptor specificity of OTase, as several interact with acceptor nascent polypeptide to influence N-glycosylation substrate selection [20–26]. In *Saccharomyces cerevisiae*, there are two isoforms of OTase defined by incorporation of either of the homologous proteins Ost3p or Ost6p [27–29], which have an ER lumenal thioredoxin-like domain [23,30] followed by four transmembrane helices. These OTase isoforms have different acceptor protein specificities [27,31,32] at the level of individual glycosylation sites [22,23].

A model of Ost3p and Ost6p function in site selection by OTase has been proposed [23], in which nascent polypeptide transiently binds to the peptide-binding groove in their thioredoxin domain, which slows protein folding and presents nearby asparagine residues to the active site of OTase for efficient glycosylation. This transient binding of stretches of nascent polypeptide is proposed to occur through noncovalent interactions and also through the formation of a transient mixed disulfide between a cysteine in

Abbreviation: O Tase, oligosaccharyl transferase.

* Corresponding author. Fax: +61 7 3365 4273.
E-mail address: b.schulz@uq.edu.au (B.L. Schulz).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.01.128

the nascent polypeptide and the active site CxxC motif of Ost3p or Ost6p [23]. Hydrophobic stretches of polypeptide do bind to the peptide-binding grooves of both Ost3p and Ost6p *in vitro*, consistent with this model [33]. The CxxC motif of Ost3p and Ost6p is also important for glycosylation of a subset of sites *in vivo* in yeast, where mutation of the Ost3p or Os6p CxxC cysteines to serines results in inefficient glycosylation at several asparagines [23]. One such affected site is Asn253 in Gas1p. Here, we performed a series of *in vitro* assays to test if Ost3p and Ost6p were capable of sequestering the model substrate polypeptide Gas1p through transient mixed disulfide bonds.

2. Materials and methods

2.1. Protein expression and purification

The malE gene was deleted from existing pMAL-c2x-OST3 and pMAL-c2x-OST6 vectors [33] and replaced with an N-terminal His10 tag by site-directed PCR mutagenesis [34] leaving the DNA sequence corresponding to the lumenal domain of Ost3p or Ost6p as predicted in UniProt Knowledgebase (position 23–185 of Ost3p/P48439/ YOR085W; position 25–188 of Ost6p/Q03723/YML019W) to give pHis10-OST3 and pHis10-OST6. TOP10 *E. coli* cells carrying pHis10-OST3 or pHis10-OST6 were grown in LB media supplemented with 100 mg/L ampicillin. Protein expression was induced at an OD600 nm of 0.7 by addition of IPTG to a final concentration of 1 mM and incubation for 4 h at 37 °C, after which the cells were harvested by centrifugation. Cells were resuspended in ice cold 50 mM potassium phosphate buffer pH 7.5, 150 mM NaCl, 10 mM imidazole, 1× complete mini EDTA free protease inhibitor cocktail (Roche) and 1 mM PMSF, and lysed with a One Shot cell disrupter (Constant Systems). Protein was purified from clarified whole cell extracts using TALON resin (Clontech) and eluted in 50 mM potassium phosphate buffer pH 7.5, 150 mM NaCl and 200 mM imidazole. Proteins were buffer exchanged with PD-10 columns (GE Healthcare) into 50 mM potassium phosphate buffer pH 7.5, 150 mM NaCl and 1 mM EDTA. *S. cerevisiae* Ost3p, Ost6p and Gas1p were expressed and purified from *E. coli* as maltose-binding protein fusions (MBP-Ost3, MBP-Ost6 and MBP-Gas1) as previously described [33]. MBP-Gas1 was eluted in 50 mM Tris–HCl buffer pH 7.5, 150 mM NaCl, 1 mM EDTA and 0.1% SDS. MBP-Ost3 and MBP-Ost6 were not eluted but were retained on amylose–agarose beads as described [33].

2.2. Whole protein mixed disulfide formation analysis

Purified MBP-Gas1 in 50 mM Tris–HCl buffer pH 7.5, 150 mM NaCl, 1 mM EDTA and 0.1% SDS was reduced by addition of DTT to a final concentration of 10 mM and incubation at 95 °C for 10 min. SDS and DTT were removed from reduced MBP-Gas1 by buffer exchange using PD-10 columns into 50 mM Tris–HCl buffer pH 7.5, 150 mM NaCl and 1 mM EDTA. In order to allow the *in vitro* formation of mixed disulfide bonds between reduced cysteines in MBP-Gas1 and the oxidized CxxC active site motif of Ost3 or Ost6, purified and reduced MBP-Gas1 was incubated with either purified Ost3, Ost6 or no additional protein in 50 mM potassium phosphate buffer pH 7.5, 150 mM NaCl and 1 mM EDTA at 25 °C for 1 h (Fig. 1). Sulfide reactivity was quenched by cysteine alkylation with addition of acrylamide to a final concentration of 25 mM and SDS to a final concentration of 0.1%, and incubation for 30 min. Protein was precipitated by addition of four volumes of 1:1 methanol:acetone, incubation at −20 °C for 16 h and centrifugation at 18,000 rcf for 10 min. The protein pellet was resuspended in 50 μL of 50 mM ammonium acetate, 10 mM DTT and 1 μg trypsin (proteomics grade, Sigma–Aldrich) and incubated at 37 °C for 16 h.

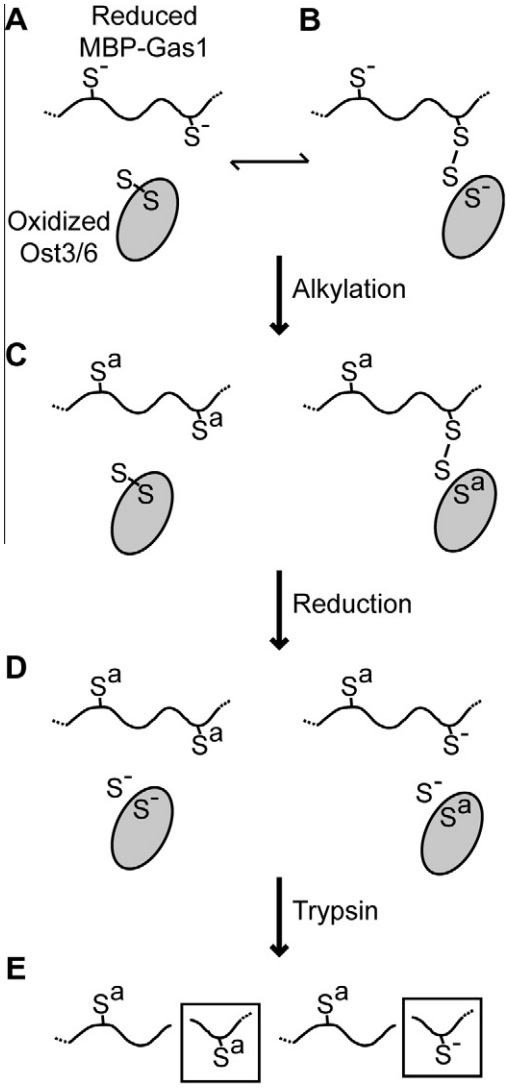


Fig. 1. Experimental design for detecting *in vitro* whole protein mixed disulfide formation with Ost3 or Ost6. (A) Reduced MBP-Gas1 is incubated together with oxidized Ost3 or Ost6. (B) A cysteine in Gas1 performs nucleophilic attack on the Ost3/6 cystine and forms a mixed disulfide bond. This reaction is reversible [23]. (C) Alkylation of free cysteines prevents further mixed disulfide shuffling. (D) Disulfide bonds are reduced with DTT. (E) Trypsin digestion and LC–ESI–MS/MS analysis allows detection of alkylated or reduced forms of the same cysteine-containing peptide (boxed). Peptides detected with an alkylated cysteine were originally free as in (A), and peptides detected with a reduced cysteine originally formed a disulfide bond as in (B). See Section 2 for details.

2.3. Peptide mixed-disulfide affinity chromatography

Purified MBP-Gas1was precipitated by addition of four volumes of 1:1 methanol:acetone and incubation at −20 °C for 16 h. Protein was resuspended in 50 μL of 50 mM ammonium acetate, 10 mM DTT and 1 μg trypsin and incubated at 37 °C for 16 h. Residual trypsin activity was inhibited by addition of PMSF to a final concentration of 1 mM and incubation at 25 °C for 30 min, followed by incubation at 95 °C for 5 min. Peptides were desalted using tC18 Sep-Pak cartridges (Waters) and eluted with 70% acetonitrile in 50 mM Tris–HCl buffer pH 7.5 with 1 mM EDTA to prevent re-oxidation of cysteines. Desalted and reduced tryptic peptides from MBP-Gas1 were applied to purified MBP-Ost3 or MBP-Ost6 still

Table 1
Identification of reduced and alkylated versions of cysteine-containing tryptic peptides from Gas1p detected by LC–ESI–MS/MS used to measure MBP-Gas1 whole protein mixed disulfide formation with Ost3 or Ost6.

Peptide sequence	*m/z*	*z*	Mass	Score
G_{48}VAYQADTANETSGSTVNDPLANYES**C**$_{74}$SR_{76}	1007.43	3	0.001	17
G_{48}VAYQADTANETSGSTVNDPLANYES<u>C</u>$_{74}$SR_{76}	1031.11	3	0.004	18
N_{253}LSIPVFFSEYG**C**$_{265}$NEVTPR$_{271}$	724.67	3	−0.001	16
N_{253}LSIPVFFSEYG<u>C</u>$_{265}$NEVTPR$_{271}$	748.35	3	−0.001	10
S_{338}YSATTSDVA**C**$_{348}$PATGK$_{353}$	779.84	2	−0.002	15
S_{338}YSATTSDVA<u>C</u>$_{348}$PATGK$_{353}$	815.36	2	−0.000	17
V_{91}YAINTTLDHSE**C**$_{103}$MK$_{105}$	575.59	3	−0.001	15
V_{91}YAINTTLDHSE<u>C</u>$_{103}$MK$_{105}$	599.27	3	−0.001	17
M_{210}TDYFA**C**$_{216}$GDDDVK$_{222}$	740.28	2	−0.003	15
M_{210}TDYFA<u>C</u>$_{216}$GDDDVK$_{222}$	775.80	2	−0.006	13
Y_{415}GAYSF**C**$_{421}$TPK$_{424}$	568.74	2	−0.003	10
Y_{415}GAYSF<u>C</u>$_{421}$TPK$_{424}$	604.26	2	−0.005	14

Reduced cysteines are bold. Alkylated cysteines are underlined. Peptide numbering is from sequence of native Gas1p.

bound to amylose-agarose beads (Fig. 3). These bead slurries were incubated in 50 mM potassium phosphate buffer pH 7.5, 150 mM NaCl and 1 mM EDTA with gentle agitation at 25 °C for 1 h. Reduced cysteines were then alkylated by addition of acrylamide to a final concentration of 25 mM and incubation for 30 min. Bead slurries were transferred to gravity flow columns and the beads were washed with 25 column volumes of 50 mM Tris–HCl buffer pH 7.5, 150 mM NaCl and 1 mM EDTA. Peptides bound to MBP-Ost3 or MBP-Ost6 were eluted with this same buffer containing in addition 10 mM DTT.

2.4. Mass spectrometry and data analysis

Peptide samples were desalted using C18 ZipTips (Milipore) prior to mass spectrometry analysis. Peptides were analysed by LC–ESI–MS/MS using a Prominence nanoLC system (Shimadzu) and TripleTof 5600 mass spectrometer with a Nanospray III interface (AB SCIEX), as previously described [35,36]. Peptides were identified using ProteinPilot (AB SCIEX), searching the LudwigNR database (downloaded from http://apcf.edu.au as at 27 January 2012; 16,818,973 sequences; 5,891,363,821 residues) with standard settings: Sample type, identification; Cysteine alkylation, acrylamide; Instrument, TripleTof 5600; Species, not limited; ID focus, biological modifications; Enzyme, trypsin; Search effort, thorough ID. False discovery rate analysis using ProteinPilot was performed on all searches. Peptides identified with greater than 99% confidence and with a local false discovery rate of less than 1% were included for further analysis, and MS/MS fragmentation spectra were manually inspected. Extracted ion chromatograms were obtained using PeakView 1.1 (AB SCIEX).

3. Results

3.1. Ost6p can form mixed disulfides with a model substrate protein

A key prediction of the proposed model of Ost3p and Ost6p function in N-glycosylation [23] is that the CxxC active sites in these proteins are capable of forming transient disulfide bonds with cysteine residues in substrate polypeptides. We tested if Ost3p and Ost6p had this activity with two complementary *in vitro* assays. For these assays we used the thioredoxin-like ER lumenal domains of Ost3p and Ost6p, and yeast Gas1p, a model glycoprotein substrate, as Asn253 in Gas1p requires the CxxC motif of either Ost3p or Ost6p for efficient glycosylation *in vivo* [22,23]. Selected peptides from Gas1p can also transiently bind non-covalently to the peptide-binding groove of both Ost3p and Ost6p *in vitro* [33]. We therefore tested if cysteine residues of Gas1p could form mixed-disulfide bonds with the active site cysteines of Ost3p and Ost6p. We expressed and purified Gas1p as a periplasmically targeted MBP-fusion protein from *E. coli*, reduced its disulfide bonds by incubation in SDS and DTT, and desalted the protein to obtain reduced MBP-Gas1 protein. We then measured the extent to which each cysteine residue was free, or alternatively involved in a disulfide bond in a cystine (Fig. 1). For this, we first alkylated free cysteines by incubation with acrylamide and then reduced the disulfide bonds with DTT. The MBP-Gas1 protein was subsequently digested to peptides with trypsin in the presence of DTT and EDTA to prevent re-oxidation of cysteines. When MBP-Gas1 protein alone was subjected to this analysis, we identified six cysteine-containing peptides from Gas1p, all of which were detected in two forms – with the cysteines either reduced or alkylated (Table 1). Reduction of MBP-Gas1 was therefore partial, mimicking incompletely oxidized polypeptide still in the process of folding, and a likely substrate of OTase. To test if Ost3 or Ost6

Table 2
LC–ESI-MS/MS identification of tryptic peptides from MBP-Gas1 in load and DTT elution fractions in peptide mixed-disulfide affinity chromatography with MBP-Ost3.

Protein	Sequence	*m/z*	*z*	ΔMass	Score
Load					
MBP	T_{371}AVINAASGR$_{380}$	480.27	2	0.001	14
MBP	G_{42}YNGLAEVGK$_{51}$	504.26	2	0.001	10
MBP	D_{322}KPLGAVALK$_{331}$	506.31	2	0.002	15
MBP	I_{343}AATMENAQK$_{352}$	538.77	2	0.002	16
MBP	T_{154}WEEIPALDK$_{163}$	601.31	2	0.000	14
MBP	V_{266}NYGVTVLPTFK$_{277}$	669.38	2	−0.001	12
MBP	E_{304}FLENYLLTDEGLEAVNK$_{321}$	699.69	3	0.003	23
MBP	F_{93}GGYAQSGLLAEITPDK$_{109}$	883.95	2	0.004	23
Gas1p	Y_{415}GAYSF**C**TPK$_{424}$	568.76	2	0.000	15
Gas1p	V_{91}YAINTTLDHSE**C**MK$_{105}$	575.60	3	−0.002	16
Gas1p	Y_{303}GLVSIDGNDVK$_{314}$	640.33	2	−0.002	18
Gas1p	M_{210}TDYFA**C**GDDDVK$_{222}$	740.29	2	−0.005	18
Gas1p	A_{106}LNDADIYVIADLAAPATSINR$_{127}$	763.07	3	0.007	27
Gas1p	S_{338}YSATTSDVA**C**PATGK$_{353}$	779.86	2	0.008	25
Gas1p	I_{194}PVGYSSNDDEDTR$_{207}$	784.35	2	0.012	17
Gas1p	T_{315}LDDFNNYSSEINK$_{328}$	830.38	2	0.002	22
Gas1p	D_{128}DPTWTVDLFNSYK$_{141}$	850.89	2	0.000	18
Gas1p	E_{468}IGSMGTNSASGSVDLGSGTESSTASSNASGSSSK$_{502}$	1065.47	3	0.002	23
Elution					
Gas1p	G_{48}VAYQADTANETSGSTVNDPLANYES**C**$_{74}$SR_{76}	1007.44	3	−0.010	25

All cysteines are reduced and bold. Peptide numbering is from sequence of native *E. coli* MBP and native *S. cerevisiae* Gas1p.

could form mixed disulfides with the cysteines in Gas1 we incubated reduced MBP-Gas1 together with either Ost3 or Ost6 and measured the extent of alkylation at each detectable cysteine in Gas1. Peptides containing the CxxC active site cysteines of Ost3p or Ost6p, which are the only cysteines in the ER lumenal domains of these proteins, were not detected in this analysis. Some cysteines, such as Cys348 in tryptic peptide Ser338-Lys353, showed no change in the extent of alkylation in the presence or absence of Ost3 or Ost6 (Fig. 2A–C). However, the extent of alkylation of some peptides did change. For example, Cys74 in tryptic peptide Gly48-Arg76 was significantly less alklyated when MBP-Gas1 had been incubated in the presence of Ost6 than when MBP-Gas1 was incubated either alone or together with Ost3 ($P < 0.05$, ANOVA) (Fig. 2D–F). This indicated that during the incubation of Ost6 and reduced MBP-Gas1, Cys74 in some MBP-Gas1 proteins was not alkylated because it formed a disulfide bond with a cysteine in the CxxC motif of Ost6 (Fig. 1B). Of the six cysteine-containing Gas1 tryptic peptides, two formed mixed disulfide bonds with Ost6 detectable with this analysis ($P < 0.05$, ANOVA) but none with Ost3 (Fig. 1G and H). This showed that the cysteines in the CxxC thioredoxin-like active site of Ost6 could form mixed disulfides with cysteines in a physiological substrate polypeptide *in vitro*.

3.2. Ost3p can form mixed disulfides with tryptic peptides from a model substrate protein

Formation of disulfides between cysteines in whole Gas1 protein and the active site cysteines of Ost3 or Ost6 may have been limited by the folding status of the partially reduced MBP-Gas1 protein. Cysteines are commonly associated with hydrophobic regions, are typically buried in folded proteins, and may be inaccessible even in a fully reduced protein. Some cysteines may therefore have not been able to form a disulfide with Ost3 or Ost6 in the context of the whole MBP-Gas1 protein (Fig. 1 and 2 and Table 1). However, interaction between substrate polypeptide and Ost3/6p *in vivo* is proposed to occur directly upon entry of the nascent polypepide from the translocon into the ER lumen [23]. To mimic this form of nascent polypeptide we tested if reduced tryptic peptides from Gas1 could form disulfide bonds with the active site cysteines of Ost3 or Ost6 (Fig. 3). We applied reduced peptides to amylose-agarose beads with pre-bound MBP-Ost3 or MBP-Ost6, alkylated free cysteines to freeze disulfide reactivity, washed away non-bound peptides and eluted peptides bound through a disulfide to Ost3/6p with DTT. We then detected peptides in the load and elution fractions using LC–ESI–MS/MS (Table 2). We detected 18 peptides from MBP-Gas1 in the load sample applied to the amylose–agarose beads with MBP-Ost3 or MBP-Ost6 (Table 2). No peptides were detected in the elution from beads with MBP-Ost6. However, one peptide, Gly48-Arg76, was detected in the DTT elution from beads with MBP-Ost3 (Table 2). This peptide contained Cys74, the same residue identified as forming a mixed disulfide with Ost6 in our whole protein mixed disulfide assay (Fig. 2D–F and H).

4. Discussion

S. cerevisiae OTase subunits Ost3p and Ost6p modulate the acceptor protein substrate range of OTase *in vivo* [22]. Both pro-

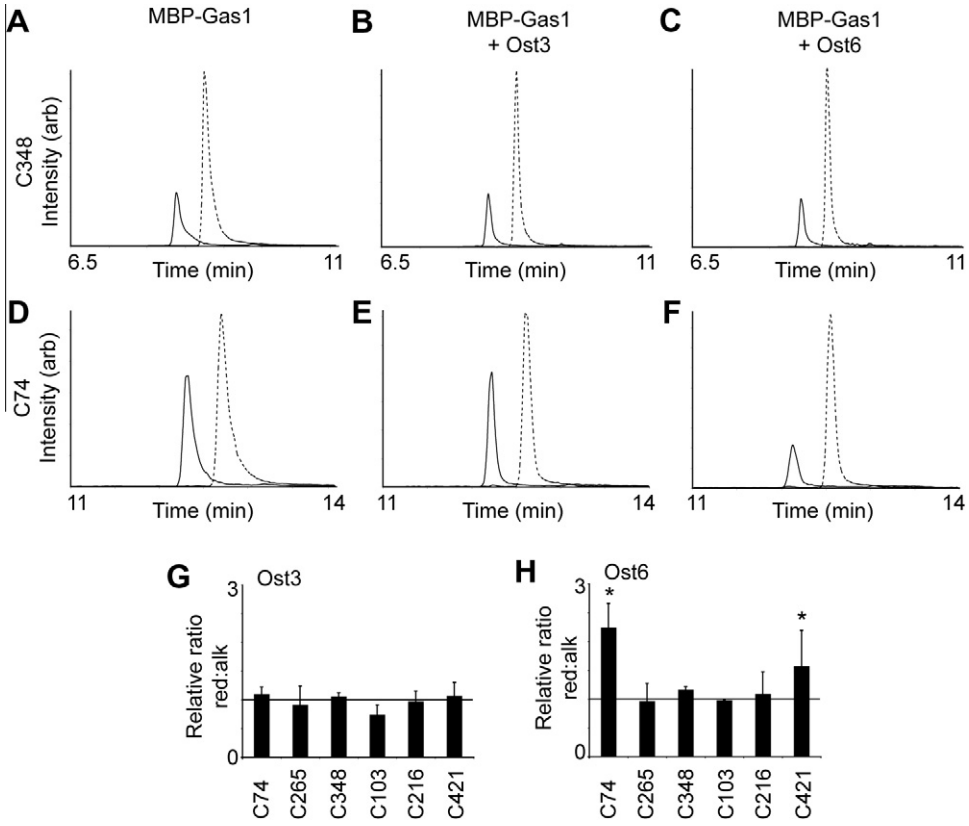


Fig. 2. *In vitro* mixed disulfide formation between MBP-Gas1 whole protein and Ost3 or Ost6. Extracted ion chromatograms of the $[M+2H]^{2+}$ ions corresponding to peptide Ser338-Lys353 with Cys348 reduced at an *m/z* of 779.84 (dashed) or alkylated at an *m/z* of 815.36 (full) from incubation of (A) MBP-Gas1 alone, (B) MBP-Gas1 with Ost3, and (C) MBP-Gas1 with Ost6. Extracted ion chromatograms of the $[M+3H]^{3+}$ ions corresponding to peptide Gly48-Arg76 with Cys74 reduced at an *m/z* of 1007.43 (dashed) or alkylated at an *m/z* of 1031.11 (full) from incubation of (D) MBP-Gas1 alone, (E) MBP-Gas1 with Ost3, and (F) MBP-Gas1 with Ost6. Ratios of the intensity of reduced and alkylated versions of peptides containing Cys74, Cys265, Cys348, Cys103, Cys216 or Cys421 from MBP-Gas1 (Gas1p amino acid numbering, Table 1) when incubated together with (G) Ost3 or (H) Ost6, relative to incubation of MBP-Gas1 alone. Values show mean of biological triplicates. Error bars show s.e.m. $^{*}P < 0.05$, ANOVA. Horizontal lines in (G) and (H) at a value of 1 are equivalent to no change.

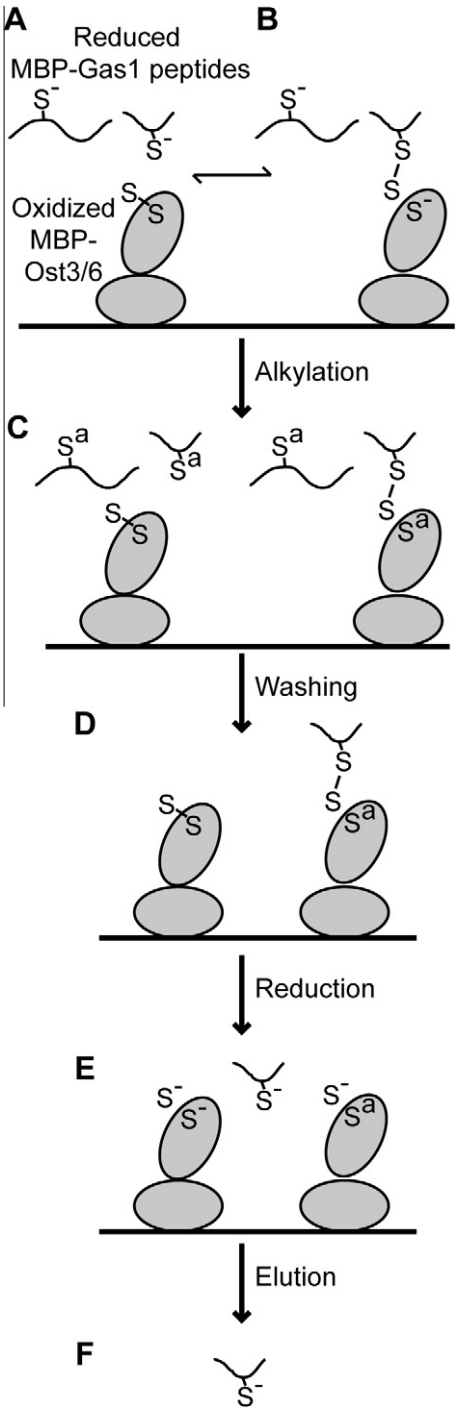


Fig. 3. Experimental design for *in vitro* peptide mixed-disulfide affinity chromatography with Ost3 or Ost6. (A) Reduced MBP-Gas1 tryptic peptides are incubated together with amylose-agarose beads with bound purified oxidized MBP-Ost3 or MBP-Ost6. (B) A cysteine in a Gas1 peptide performs nucleophilic attack on the Ost3/6 cystine and forms a mixed disulfide bond. This reaction is reversible [23]. (C) Alkylation of free cysteines prevents further mixed disulfide shuffling. (D) Non-bound peptides are removed by thorough washing. (E) Reduction with DTT releases peptides bound to Ost3 or Ost6 through a disulfide bond. (F) LC–ESI–MS/MS analysis allows detection of peptides. See Materials and Methods for details.

teins have an ER lumenal thioredoxin-like domain, whose CxxC active site residues are required for efficient *in vivo* glycosylation of a subset of glycosylation sites, including N253 in Gas1p [23]. It has been proposed that the active site cysteines of Ost3p and Ost6p can form transient mixed disulfides with nascent polypeptide substrate [23]. To test this, we developed two independent *in vitro* assays to identify cysteines in *S. cerevisiae* Gas1p that could form transient mixed disulfide bonds with the thioredoxin-like active site cysteines of Ost3p or Ost6p. One of these assays tested interactions using Gas1p as a reduced but intact MBP fusion protein (Fig. 1), and the other used this same protein after it had been digested to peptides with trypsin (Fig. 3). The combination of these approaches identified that the Cys74 cysteine residue in Gas1p could form a mixed disulfide bond with both Ost3p and Ost6p (Fig. 2, Table 2). In addition, the whole protein interaction approach identified that Cys421 could form a mixed disulfide bond with Ost6p (Fig. 2). While efficient glycosylation of Asn253 in Gas1p *in vivo* requires the CxxC motif of either Ost3p or Ost6p [23], it is not clear if this requirement is due to the formation of a mixed disulfide bond between Ost3p or Ost6p and either Cys74 or Cys421 in Gas1p, that our analyses here detected could bind *in vitro*. Both Cys74 and Cys421 are distant from Asn253 both in amino acid sequence and in the predicted final folded protein structure of Gas1p [37,38]. It is not clear if the peptides that we detected forming mixed disulfides with Ost3p and Ost6p *in vitro* also form such transient bonds *in vivo*, or if these particular bonds are key for glycosylation at Asn253 of Gas1p. Although we detected six peptides from Gas1p that contain cysteine residues (Table 1), there are eight additional cysteine residues in Gas1p that we did not detect with our analytic approach. Some of these cysteine residues may also form mixed disulfide bonds with Ost3p and/or Ost6p relevant to N-glycosylation of Asn residues in Gas1p *in vivo*.

In summary, our results validated a key prediction of the proposed model of Ost3/6p function in N-glycosylation [23]. It has been previously shown that the CxxC motifs of Ost3p and Ost6p are important for glycosylation *in vivo* [23]. Here, we showed that the ER lumenal domains of both Ost3p and Ost6p could form mixed disulfides with cysteines in their physiological protein substrate Gas1p, using a peptide and a whole-protein assay. Ost3p and Ost6p showed differences in their activities in these assays, consistent with their different peptide-binding activities *in vitro* [33] and different range of substrates *in vivo* [22]. The assays we developed may also be applicable to *in vitro* identification of transient mixed disulfide bonds in other systems.

Acknowledgments

This work was supported by NHMRC Project Grant 631615 and NHMRC CDF APP1031542 to B.L.S.

References

[1] K. Ohtsubo, J.D. Marth, Glycosylation in cellular mechanisms of health and disease, Cell 126 (2006) 855–867.
[2] D. Calo, L. Kaminski, J. Eichler, Protein glycosylation in archaea: sweet and extreme, Glycobiology 20 (2010) 1065–1076.
[3] H. Nothaft, C.M. Szymanski, Protein glycosylation in bacteria: sweeter than ever, Nat. Rev. Microbiol. 8 (2010) 765–778.
[4] Y. Harada, H. Li, H. Li, W.J. Lennarz, Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site, Proc. Natl. Acad. Sci. USA 106 (2009) 6945–6949.
[5] D.J. Kelleher, R. Gilmore, An evolving view of the eukaryotic oligosaccharyltransferase, Glycobiology 16 (2006) 47R–62R.
[6] F. Schwarz, M. Aebi, Mechanisms and principles of N-linked protein glycosylation, Curr. Opin. Struct. Biol. 21 (2011) 576–582.
[7] J.D. Oliver, H.L. Roderick, D.H. Llewellyn, S. High, ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin, Mol. Biol. Cell 10 (1999) 2573–2582.
[8] R. Daniels, B. Kurowski, A.E. Johnson, D.N. Hebert, N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin, Mol. Cell 11 (2003) 79–90.
[9] A. Helenius, M. Aebi, Roles of N-linked glycans in the endoplasmic reticulum, Annu. Rev. Biochem. 73 (2004) 1019–1049.
[10] B.R. Pearse, D.N. Hebert, Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum, Biochem. Biophys. Acta 2010 (1803) 684–693.
[11] A. Varki, Biological roles of oligosaccharides: all of the theories are correct, Glycobiology 3 (1993) 97–130.

[12] B.L. Schulz, A.J. Sloane, L.J. Robinson, S.S. Prasad, R.A. Lindner, M. Robinson, P.T. Bye, D.W. Nielson, J.L. Harry, N.H. Packer, N.G. Karlsson, Glycosylation of sputum mucins is altered in cystic fibrosis patients, Glycobiology 17 (2007) 698–712.

[13] B.L. Schulz, A.J. Sloane, L.J. Robinson, L.T. Sebastian, A.R. Glanville, Y. Song, A.S. Verkman, J.L. Harry, N.H. Packer, N.G. Karlsson, Mucin glycosylation changes in cystic fibrosis lung disease are not manifest in submucosal gland secretions, Biochem. J. 387 (2005) 911–919.

[14] B.L. Schulz, J.C. White, C. Punyadeera, Saliva proteome research: current status and future outlook, Crit. Rev. Biotechnol. (2012).

[15] F. Dall'Olio, N. Malagolini, M. Trinchera, M. Chiricolo, Mechanisms of cancer-associated glycosylation changes, Front Biosci. 1 (2012) 670–699.

[16] E. Mohorko, R. Glockshuber, M. Aebi, Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation, J. Inherit. Metab. Dis. 34 (2011) 869–878.

[17] P. Roboti, S. High, Keratinocyte-associated protein 2 is a bona fide subunit of the mammalian oligosaccharyltransferase, J. Cell Sci. 125 (2012) 220–232.

[18] J.H. Lee, W.H. Yu, A. Kumar, S. Lee, P.S. Mohan, C.M. Peterhoff, D.M. Wolfe, M. Martinez-Vicente, A.C. Massey, G. Sovak, Y. Uchiyama, D. Westaway, A.M. Cuervo, R.A. Nixon, Lysosomal proteolysis and autophagy require presenilin 1 and are disrupted by Alzheimer-Related PS1 mutations, Cell 141 (2010) 1146–1158.

[19] C. Lizak, S. Gerber, S. Numao, M. Aebi, K.P. Locher, X-ray structure of a bacterial oligosaccharyltransferase, Nature 474 (2011) 350–355.

[20] B.L. Schulz, Beyond the sequon: sites of N-glycosylation, in: S. Petrescu (Ed.), Glycosylation, Intech, 2012.

[21] P. Roboti, S. High, The oligosaccharyltransferase subunits OST48, DAD1 and KCP2 function as ubiquitous and selective modulators of mammalian N-glycosylation, J. Cell Sci. 125 (2012) 3474–3484.

[22] B.L. Schulz, M. Aebi, Analysis of glycosylation site occupancy reveals a role for Ost3p and Ost6p in site-specific N-Glycosylation efficiency, Mol. Cell. Proteomics 8 (2009) 357–364.

[23] B.L. Schulz, C.U. Stirnimann, J.P.A. Grimshaw, M.S. Brozzo, F. Fritsch, E. Mohorko, G. Capitani, R. Glockshuber, M.G. Grütter, M. Aebi, Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency, Proc. Natl. Acad. Sci. USA 106 (2009) 11061–11066.

[24] C.M. Wilson, S. High, Ribophorin I acts as a substrate-specific facilitator of N-glycosylation, J. Cell Sci. 120 (2007) 648–657.

[25] C.M. Wilson, C. Kraft, C. Duggan, N. Ismail, S.G. Crawshaw, S. High, Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon, J. Biol. Chem. 280 (2005) 4195–4206.

[26] C.M. Wilson, Q. Roebuck, S. High, Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core, Proc. Natl. Acad. Sci. USA 105 (2008) 9534–9539.

[27] M. Schwarz, M. Knauer, L. Lehle, Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit, FEBS Lett. 579 (2005) 6564–6568.

[28] U. Spirig, D. Bodmer, M. Wacker, P. Burda, M. Aebi, The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast, Glycobiology 15 (2005) 1396–1406.

[29] A. Yan, W.J. Lennarz, Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons, Glycobiology 15 (2005) 1407–1415.

[30] J.S. Fetrow, N. Siew, J.A. Di Gennaro, M. Martinez-Yamout, H.J. Dyson, J. Skolnick, Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight?, Protein Sci 10 (2001) 1005–1014.

[31] D. Karaoglu, D.J. Kelleher, R. Gilmore, Functional characterization of Ost3p. Loss of the 34-kD subunit of the Saccharomyces cerevisiae oligosaccharyltransferase results in biased underglycosylation of acceptor substrates, J. Cell Biol. 130 (1995) 567–577.

[32] R. Knauer, L. Lehle, The oligosaccharyltransferase complex from Saccharomyces cerevisiae. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex, J. Biol. Chem. 274 (1999) 17249–17256.

[33] M.F.B. Jamaluddin, U.M. Bailey, N.Y.J. Tan, A.P. Stark, B.L. Schulz, Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p, Protein Sci. 20 (2011) 849–855.

[34] Y. Imai, Y. Matsushima, T. Sugimura, M. Terada, A simple and rapid method for generating a deletion by PCR, Nucleic Acids Res. 19 (1991) 2785.

[35] U.M. Bailey, M.F.B. Jamaluddin, B.L. Schulz, Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins, J. Proteome Res. 11 (2012) 5376–5383.

[36] U.M. Bailey, C. Punyadeera, J.J. Cooper-White, B.L. Schulz, Analysis of the extreme diversity of salivary alpha-amylase isoforms generated by physiological proteolysis using liquid chromatography-tandem mass spectrometry, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 911 (2012) 21–26.

[37] R. Hurtado-Guerrero, A.W. Schüttelkopf, I. Mouyna, A.F. Ibrahim, S. Shepherd, T. Fontaine, J.P. Latgè, D.M. van Aalten, Molecular mechanisms of yeast cell wall glucan remodeling, J. Biol. Chem. 284 (2009) 8461–8469.

[38] L. Popolo, E. Ragni, C. Carotti, O. Palomares, R. Aardema, J.W. Back, H.L. Dekker, L.J. de Koning, L. de Jong, C.G. de Koster, Disulfide bond structure and domain organization of yeast beta(1,3)-glucanosyltransferases involved in cell wall biogenesis, J. Biol. Chem. 283 (2008) 18553–18565.